

STIC-Biotech/ChemLib

66971

From: Ton, Thaia
Sent: Monday, May 20, 2002 9:41 AM
To: STIC-Biotech/ChemLib
Cc: Ton, Thaia
Subject: Sequence Search Request

I would like to request a search for Application No. 09/696,686 [Inventor: Allen, Filing Date: 10/26/2000]

A standard sequence search for SEQ ID NOS: 45, 46, 47, 50 and 51 against the commercial nucleic acid databases.

Thank you very much.

Thaia N. Ton

Patent Examiner

Art Unit 1632

Room 12A16 CM1

Mailbox: 12E12 CM1

(703) 305-1019

RECEIVED

of Contact:
Sheppard

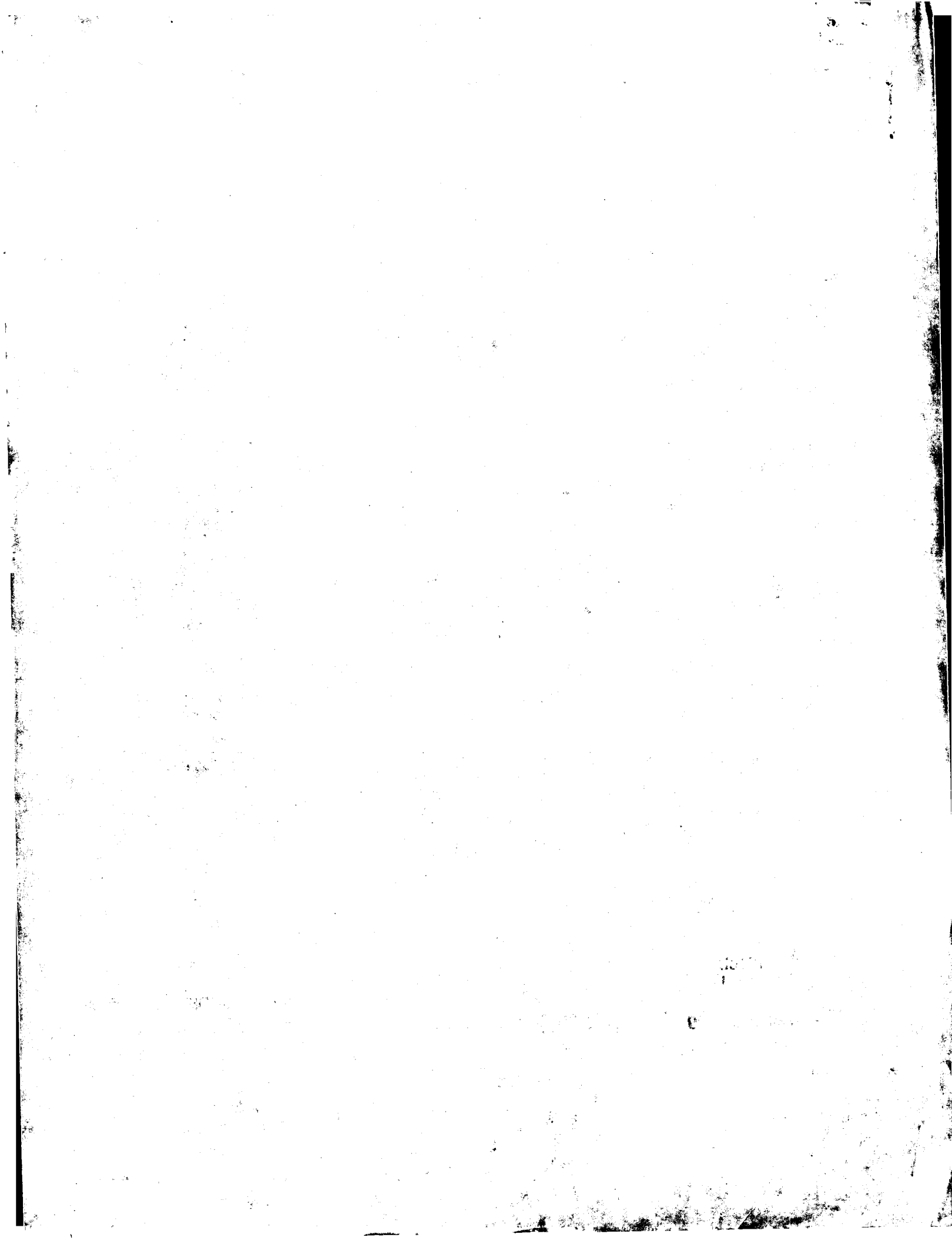
Searcher:
Phone:
Location:
Date Picked Up:
Date Completed: 5/25/02
Searcher Prep/Review:
Clerical:
Online time:

TYPE OF SEARCH:

NA Sequences:
AA Sequences:
Structures:
Bibliographic:
Litigation:
Full text:
Patent Family:
Other:

VENDOR/COST(where applic.)

STN:
DIALOG:
Questel/Orbit:
DRLink:
Lexis/Nexis:
Sequence Sys.:
WWW/Internet:
Other (specify):



Ton, Thaia

From: Ton, Thaia
Sent: Monday, May 20, 2002 9:41 AM
To: STIC-Biotech/ChemLib
Cc: Ton, Thaia
Subject: Sequence Search Request

I would like to request a search for Application No. 09/696,686 [Inventor: Allen, Filing Date: 10/26/2000]

A standard sequence search for SEQ ID NOS: 45, 46, 47, 50 and 51 against the commercial nucleic acid databases.

Thank you very much.

Thaia N. Ton

Patent Examiner

Art Unit 1632

Room 12A16 CM1

Mailbox: 12E12 CM1

(703) 305-1019



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 23:06:04 ; Search time 4203.75 seconds
(without alignments)
1187.957 Million cell updates/sec

Title: US-09-696-686-51

Perfect score: 370

Sequence: 1 tgcgattgccagcaaatgc.....atgaggacgcctgcggcc 370

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: gb_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.2	13.8	1090	12	AG128386 Pan trogl
C 2	50.6	13.7	941	12	CNS06LUR7
C 3	49.8	13.5	1101	12	CNS016MU
4	49.6	13.4	723	12	AZ184008 Drosophil
5	49.4	13.4	1196	10	BF256755 SP-1002_A
6	49.2	13.3	882	12	BF256755 HVSMEF001
7	49	13.2	660	12	AG126129 Pan trogl
C 8	48.6	13.1	1101	12	AZ198637 SP_1037_B
C 9	48.4	13.1	295	12	CNS006GJM
C 10	48.2	13.0	458	10	CNS00EVS
11	48	13.0	843	12	BE348494 ht69g12_x
C 12	47.8	12.9	466	10	AG126129 Pan trogl
C 13	47.8	12.9	937	10	AG126129 Pan trogl
C 14	47.8	12.9	1101	12	AG126129 Pan trogl
C 15	47.6	12.9	922	12	CNS00G7F
C 16	47.4	12.8	648	12	BH021826 GH.MBB000
17	47.4	12.8	762	12	AG127831 Pan trogl

18	47.4	12.8	918	12	AQ895329 HS_4832_A
19	47.4	12.8	933	12	CNS04125 Tetraodon
C 20	47.4	12.8	1638	10	BG179554 602348165
C 21	47.2	12.8	271	12	CNS008FE
C 22	47.2	12.8	544	12	CNS015XA
C 23	47.2	12.8	780	12	B21793
C 24	47.2	12.8	996	12	CNS04OW3
C 25	47.2	12.8	999	12	CNS02ITK
26	47.2	12.8	1026	12	AG081192 Tetraodon
27	47.2	12.8	1293	10	BF256498 HVSMEF001
28	47	12.7	810	12	AG127829 Pan trogl
29	47	12.7	855	9	AL572700
30	47	12.7	927	12	CNS06XJF
C 31	47	12.7	964	12	CNS003WG
C 32	46.8	12.6	378	9	AL535833
C 33	46.8	12.6	913	12	AG076623
C 34	46.8	12.6	936	12	CNS01608
C 35	46.8	12.6	999	12	AG132063
C 36	46.8	12.6	1235	10	BM475329 AGENCOURT
37	46.6	12.6	511	9	AL539262
38	46.6	12.6	684	10	BG786306 SEAMC006
39	46.6	12.6	741	12	AG049269 Pan trogl
C 40	46.6	12.6	874	12	AZ207852 SP_0135_A
41	46.6	12.6	927	12	AG099877 Pan trogl
42	46.4	12.5	465	12	BH254587 SALK_0168
43	46.4	12.5	919	12	AG132186 Pan trogl
44	46.4	12.5	980	10	BG786274 SEAMC006
45	46.2	12.5	812	9	AL537183

ALIGNMENTS

RESULT 1

AG128386
LOCUS Pan troglodytes DNA, clone: PTB-139113.R, genomic survey sequence.
DEFINITION AG128386
ACCESSION AG128386
VERSION AG128386.1 GI:16657551
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
ORGANISM Pan troglodytes

REFERENCE

AUTHORS 1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1090)

AUTHORS

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

COMMENT

PRIMERS

Sequencing: M13rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 1090

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-139113.R"

FEATURES

source

REFERENCE
AUTHORS

/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEF0010017f"
/clone_lib="Hordeum vulgare seedling root EST library
HVCNA0007 (Etiolated and unstressed)"
/issue_type="Seedling root"
/lab_host="TJ121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedling roots were
then harvested, total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give pBluescript
SK(-) cDNA phagemids. These steps were performed in the TJ
Close laboratory at the University of California,
Riverside (Choi, Close, Fenton). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders. Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/gpages/bgn/31/cover.html)."

TITLE
JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbescgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB this BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

COMMENT

PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1..882
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-136J13.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 0 a 1 c 536 g 196 t 149 others
ORIGIN

FEATURES
source

Query Match 13.4%; Score 49.4; DB 10; Length 1196;
Best Local Similarity 47.8%; Pred. No. 0.77;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

BASE COUNT

75 a 35 c 948 g 105 t 33 others
ORIGIN

Query Match

17 atgcgaaggtgagggggcgccgctgagcagacctggcctgctctctctctctctctcgcctgc 136
Db 650 AGGGGGGGGTGCTGTTGG 709
Qy 77 gggccctgtggaaggtctggcctgagcagacctggcctgctctctctctctctctcgcctgc 136
Db 710 GGGGGGGGGTGG 769
Qy 137 tgtgtcttgcgcgaatgctgggtgctgtgactctcgataaacttgagatccctgctt 196
Db 770 TGGGTGGGTGGGGGTGCTGCTGGGGGGGGTGGGGGGGGTGGGGGGGTGGGGGGGTG 829
Qy 197 ttggcgcaatccggggtagttgctcatcaagactagagtggggtggaggggaagcctt 256
Db 830 GTGTGGGGGTGGGA 889
Qy 257 catcacgaagcctctgcgaatgaagagttggccaggggaaagcattgctgagagg 315
Db 890 GGGACGGGGGGGGGGGGGGTGGGGGGGGTGGGGAGGTGGAGGGGGGGGGGGGGGGGG 948

RESULT

6
AG126129 882 bp DNA linear GSS 04-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-136J13.F; genomic survey sequence.
DEFINITION AG126129
ACCESSION AG126129
VERSION GSS: GSS (genome survey sequence).
KEYWORDS Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
SOURCE BAC Library clone:PTB-136J13.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 (sites)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 882)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbescgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB this BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

TITLE

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbescgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB this BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbescgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB this BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

COMMENT

PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1..882
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-136J13.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 0 a 1 c 536 g 196 t 149 others
ORIGIN

FEATURES

Query Match 13.3%; Score 49.2; DB 12; Length 882;
Best Local Similarity 47.9%; Pred. No. 0.78;
Matches 114; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

BASE COUNT

19 gcgaaggtgagggggcgccgctgagcagacctggcctgctctctctctctctcgcctgc 78
Db 155 GGGGGGGGGGNNNGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 214
Qy 79 ggcctgtggaaggtctggtgctggcagacctggcctgctctctctctctctcgcctgc 138
Db 215 GGG 274
Qy 139 tgtgtcttgcgcgaatgctgggtgctgtgactctcgataaacttgagatccctgctt 198
Db 275 TGGTGTGGTGTG 334
Qy 199 ggcgaatccggggtagttgctcatcaagactagagtggggtggaggggaagcctt 256
Db 335 TGGGGGTGGGGGGGGTGTGTTGGGGTGGGGGGGGGGGGGGTGTGTTGGTGGTT 392

RESULT

7
AZ198637 660 bp DNA linear GSS 31-AUG-2000
LOCUS SP_1037_B2_F11.SP6E Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library Strongylocentrotus purpuratus
genomic clone Plate-1037 Col-22 Row-L, DNA sequence.
DEFINITION AZ198637
ACCESSION AZ198637
VERSION GSS:
KEYWORDS Strongylocentrotus purpuratus.
SOURCE Strongylocentrotus purpuratus
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoida; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 660)
Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray

and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
SOURCE

```
1. .295
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCL-98"
/clone="BACR29P06"
note="end : T7"
```

BASE COUNT	41 a	200 c	5 g	26 t	23 others
ORIGIN					

Query Match	13.1%;	Score 48.4;	DB 12;	Length 295;
Best Local Similarity	45.6%;	Pred. NO. 0.88;		
Indels	116;	Mismatches	14;	Gaps 0;

Qy	19	gcgaagctgaggggcgggccgagggcgtacccaagcccagggcgagggcgagggggggcggg	78
Db	257	GGGCGGGGGGGAGGG	198
Qy	79	ggcctgtggaaagggctctgggacctggcagaccctggctggggtcctcttgggccctgctg	138
Db	197	GGGTGTGKGGGGGGGATAGKGGGGKGGGGGGGKKGKGGKGGTTTAAAGKGGGGKG	138
Qy	139	tgtgcttgccgcaactcgtgggtgctgtgacttcgataaacctcgagatcacctcgtttt	198
Db	137	TGTAGATGGGAGAGCTGKGGGGGGGTGTGGGGTKKGGSGTGKGGTGGGGG	78
Qy	199	ggcgcaatccgggggttagtttgtctcatcaagactagagggtgggggttagagggaagcttc	257
Db	77	GSGAGCTFRGGGATGGGTGAATAATGATCANACGTGATCTGGGTGGGAGGAGTGTCC	19

RESULT 10
BE348494/c
LOCUS BE348494 458 bp mRNA linear EST 18-JUL-2000
DEFINITION ht59g12.x1 NC1_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3152038 3,
similar to contains element MSRL repetitive element ; mRNA
sequence.

BE348494
 BE348494.1 GI:9260347
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 458)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapps-r@mail.nih.gov
 Tissue Procurement: Christopher Mokaluk, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 CDNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
 Seq primer: -400P from gibco
 High quality sequence stop: 457.

```

FEATURES
source
Location/Qualifiers
1. .458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:3152038"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"

```

/note="Organ: lung; Vector: pTTT30-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520304-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT	0 a	313 c	22 g	0 t	123 others
ORIGIN					

[illegible]

RESULT	11
AQ38512	
TACUS	
DEFINITION	AQ738512 linear DNA 843 bp
DESCRIPTION	HS_5381_B1_L1_TTA RPCI-11 Human Male BAC library Homo sapiens genomic clone Plate=957 Col=21 Row=D, DNA sequence.
ACCESSION	AQ738512
VERSION	AQ738512.1 GI:5516034
KEYWORDS	GSS.

SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 843)
Mahairis,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,S., Zhao,S., Adams,M.D. and
Hood,L.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (<http://www.htsc.washington.edu>). BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 957 row: D column: 21
Seq primer: T7

Fri May 24 10:49:52 2002

Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45
 Email: Fred.Aspegren@kpat.slu.se
 Seq primer: T7 primer.
 Location/Qualifiers
 1. 937
 /organism="Pinus sylvestris/Heterobasidion annosum"
 /db_xref="taxon:169015"
 /clone="hasp001xa23f"
 /clone_lib="Heterobasidion annosum - Scots pine infection
 stage subtraction cDNA library (hasp)"
 /dev_stage="Seedling roots of scots pine were infected for
 6 days with H. annosum"
 /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
 hybridization cDNA library was constructed from scots
 pine roots infected for 6-days with mycelia of
 Heterobasidion annosum (FP5)."
 767 21 2 others

BASE COUNT	102 a	45 c	767 g	21 t	2 others
ORIGIN					
Query Match	12.9%	Score	47.8;	DB	10; Length 937;
Best Local Similarity	47.8%;	Pred. No.	1.6;		
Matches 139; Conservative	0;	Mismatches	152;	Indels	0; Gaps
QY	25	tgtaggggcccggccggcgctagccaagcccgagggcgggcgagagggcgggcgccctg	84		
Db	373	GGGTGGGGGGGGGGGAGGTGGTGGAGAGCGGGGAGGGGAGGGGGGGCGGG	432		
QY	85	tyggaaagggctcggccctggcagaacctggctggggtctcttgccctctgtgtgct	144		
Db	433	CGCGAGGCGGTGGCGTGGCGAGGGTGGAGGGGGGGGGGTAGAGCGCGGGGGGAGGAG	492		
QY	145	ttagcgcaatgctgggtgtgtgactctctggataacctggagatccctgtctttgggga	204		
Db	493	GGGGAGGGGAGGGTGGCGAGGGCGGTAGGGGGGGGGGAGGGCGGGCAGGGGGCGG	552		
QY	205	atcccggggtagttgtctcatcaagactagagtgggggtggagggaaggcttcatacag	264		
Db	553	AGGGGGGGGAGGGTGGGGAGGGGAGGGAGGGGGGGGGGAGGGGGGGGGAGGGGGG	612		
QY	265	aagcctgtctgcaaatcgaagagttggccaaggaagaacatggcgctgcagagg	315		
Db	613	GAGGGGGGGGGGGGGCGAGGGGGGGGGGGGGGGGGGGGGAGGAGGTAGAGGGG	563		

RESULT	14
CNS00G7F/c	
LOCUS	
DEFINITION	CNS00G7F linear GSS 03-JUN-1999 Drosophila melanogaster genome survey sequence 17 end of BAC: BACR32M17 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL071731
VERSION	AL071731.1 GI:4951571
KEYWORDS	GSS
SOURCE	fruit fly. <i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
ORGANISM	1 (bases 1 to 1101) Genoscope.
REFERENCE	Direct Submission
AUTHORS	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
TITLE	Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL	- Web : www.genoscope.cns.fr)

COMMENT

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named pPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the library p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	Source	Location/Qualifiers
	1. ll101	
		organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone_lib="RPCI-98"
		/clone="BACR32M17"
		/note="end : T7"
176 a	481 c	176 g
176 a	481 c	155 t
176 a	481 c	113 others

BASE COUNT	176 a	481 c	176 g	155 t	113 others
ORIGIN					
Query Match		12.9%	Score 47.8;	DB 12;	Length 1101;
Best Local Similarity		43.4%	Pred. No. 1.7;		

[illegible]

RESULT	15
CNS06WDK	
LOCUS	922 bp DNA linear
DEFINITION	T7 end of clone AX0AA030F01 of library AX0AA from strain CBS 7064 of Pichia farinosa, genomic survey sequence.
	GSS 06-JUL-2001

ACCESSION	AL418302	1	GI:12200893
VERSION	AL418302.1		
KEYWORDS	GSS.		
SOURCE	Pichia farinosa.		
ORGANISM	Pichia farinosa		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
AUTHORS	Saccharomycetales; Saccharomycetaceae; Pichia.		
TITLE	1 (bases 1 to 922) de Montigny, J., Spelner, C., Souciet, J., Tekai, F., Dujon, B., Wincker, P., Artiguenave, F. and Potier, S. Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila FEBS Lett. 487 (1), 87-90 (2000)		
JOURNAL	2 (bases 1 to 922) Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Liorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekai, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies FEBS Lett. 487 (1), 3-12 (2000)		
MEDLINE	20584725		
REFERENCE	3 (bases 1 to 922) Genoscope. Direct Submission		
AUTHORS			
TITLE			
JOURNAL			
MEDLINE	20584711		
REFERENCE	3 (bases 1 to 922) Genoscope. Direct Submission		
AUTHORS			
TITLE			

COMMENT

Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremlieu, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

```

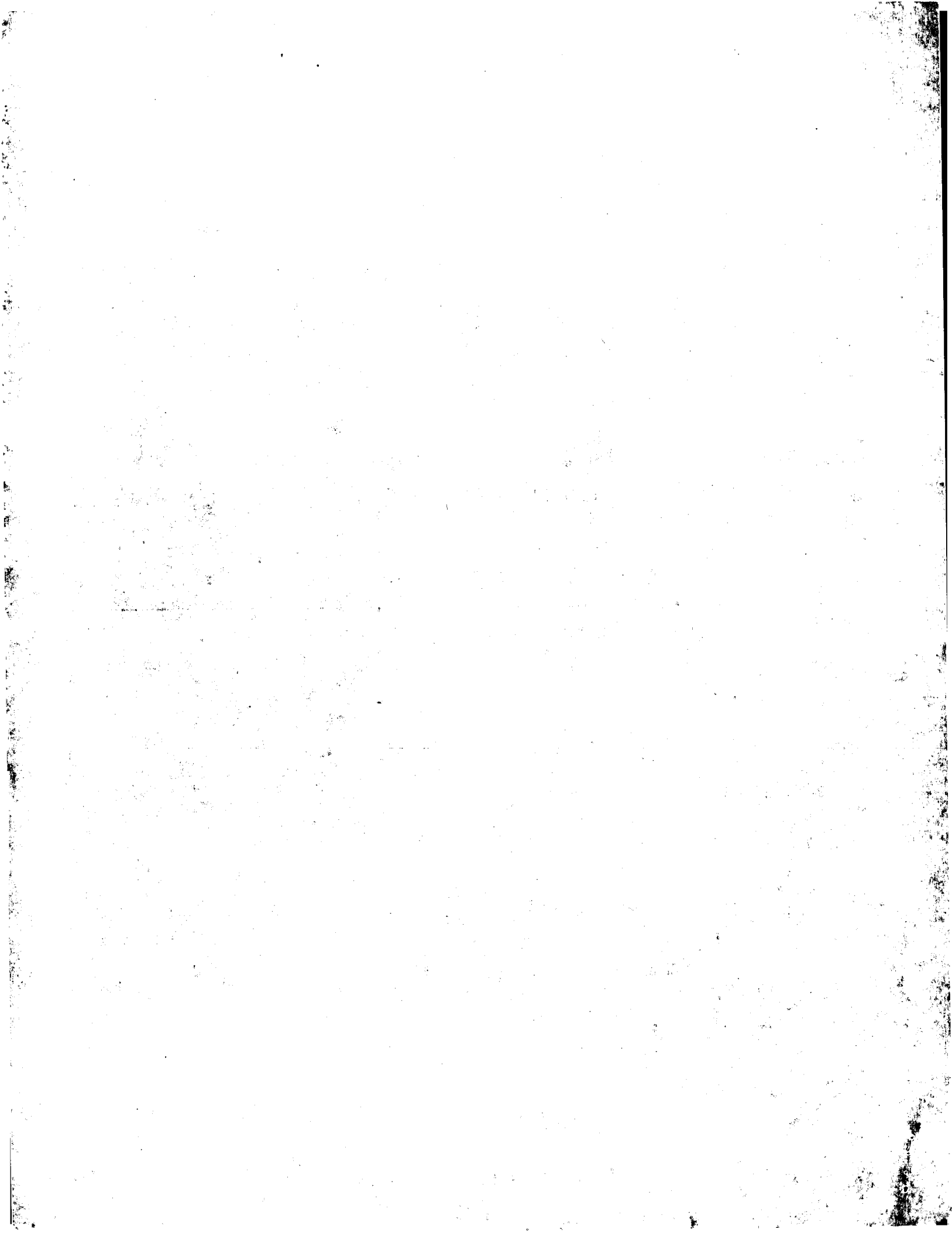
FEATURES
  source
    Location/Qualifiers
      1..922
        /organism="Pichia farinosa"
        /strain="CBS 7064"
        /db_xref="taxon:4920"
        /clone="AX0AA030F01"
        /clone_lib="AX0AA"
        /note="end: 77"

```

BASE COUNT	216 a	98 c	376 g	138 t	94 others
ORIGIN					

Query Match	12.9%	Score 47.6;	DB 12;	Length 922;
Best Local Similarity	42.4%;	Pred. No. 1.7;	Mismatches 14;	Gaps 0;
Matches 104;	Conservative	Indels	Indels	Gaps
QY	19	gcgaaggtgaggggcgggccgcgagcctagccaagcccagggcgaggagggcgcg	78	
Db	327	GCNAGGGGSKSGGGGGGGGGCGGGRGGGGGGGGGGGGGGGGGGGGGGGGGGG	386	
QY	79	ggcctgtggaaaggctggggcctgycgagacactgggctggggtctctcttggccctgctg	138	
Db	387	GGG	446	
QY	139	tgtgcttttcgcgaagtctggtgctgtgactctcgataaacctcgagatccctgctttt	198	
Db	447	GGG	506	
QY	199	ggcgcaatccgggggtagtgtctcatcaagactagagtggtgggggtggagggaaggtcca	258	
Db	507	GGGKSGGGGGGGGGRRAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGWA	566	
QY	259	tacag	263	
Db	567	TATWG	571	

Search completed: May 23, 2002, 23:06:13
Job time: 7529 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:20:45 ; Search time 121.39 seconds
(without alignments)
748.698 Million cell updates/sec

Title: US-09-696-686-51

Perfect score: 370

Sequence: 1 tgcgattcccgcaaatgc.....atgagcgccctgcggcc 370

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	50.4	13.6	7218	1	US-08-232-463-14
C 2	39.2	10.6	319	4	US-09-165-264-8
C 3	39	10.5	152331	3	US-09-128-155-16
C 4	37.6	10.2	318	4	US-09-165-264-12
C 5	37.6	10.2	320	4	US-09-165-264-13
C 6	37.2	10.1	320	4	US-09-165-264-14
C 7	37	10.0	320	4	US-09-165-264-7
C 8	36	9.7	4405	1	US-07-885-972A-3
C 9	36	9.7	4405	2	US-08-745-880-3
C 10	36	9.7	4405	2	US-08-480-382-3
C 11	35.8	9.7	4488	4	US-08-406-030A-3
C 12	35.6	9.6	320	4	US-09-165-264-11
C 13	35.4	9.6	4046	1	US-07-793-961A-1
C 14	35.4	9.6	4046	1	US-08-240-357-1
C 15	34.8	9.4	374	2	US-08-370-156-24
C 16	34.8	9.4	1215	2	US-08-370-156-26
C 17	34.8	9.4	2256	2	US-08-318-826A-5
C 18	34.8	9.4	2256	2	US-08-370-156-1
C 19	34.8	9.4	2256	3	US-08-814-095-1
C 20	34.8	9.4	3016	2	US-08-318-826A-7
C 21	34.8	9.4	3016	2	US-08-370-156-5
C 22	34.8	9.4	3016	3	US-08-814-095-5
C 23	34.8	9.4	3096	2	US-08-318-826A-6
C 24	34.8	9.4	3096	2	US-08-370-156-3
C 25	34.8	9.4	3096	3	US-08-814-095-3
C 26	34.2	9.2	53525	3	US-08-658-136-2
C 27	34.2	9.2	53577	3	US-08-658-136-1

28	34	9.2	1737	2	US-08-750-703-2	Sequence 2, Appli
29	34	9.2	8779	2	US-08-750-703-4	Sequence 4, Appli
C 30	33.8	9.1	2689	2	US-08-985-090-1	Sequence 1, Appli
C 31	33.8	9.1	2689	3	US-09-165-543-1	Sequence 11, Appli
C 32	33.6	9.1	2663	4	US-09-499-884-11	Patent No. 5215881
C 33	33.4	9.0	2721	6	5215881-2	Sequence 1, Appli
C 34	33.4	9.0	4257	2	US-08-690-473-1	Sequence 1, Appli
C 35	33.4	9.0	4257	4	US-09-259-821A-1	Sequence 1, Appli
C 36	33.4	9.0	4257	4	US-08-843-659-1	Sequence 1, Appli
C 37	33.4	9.0	4425	2	US-08-749-169A-1	Sequence 1, Appli
C 38	33.4	9.0	4425	2	US-09-130-032A-1	Sequence 1, Appli
C 39	33.4	9.0	12001	1	US-08-458-588A-11	Sequence 11, Appli
40	33	8.9	1702	1	US-08-459-586-9	Sequence 9, Appli
41	33	8.9	1702	2	US-08-282-696-9	Sequence 9, Appli
42	33	8.9	8438	1	US-07-945-283-1	Sequence 1, Appli
43	33	8.9	18994	1	US-08-459-586-4	Sequence 4, Appli
44	33	8.9	18994	2	US-08-282-696-4	Sequence 4, Appli
45	32.8	8.9	640	2	US-08-835-099A-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)836-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match 13.6%; Score 50.4; DB 1; Length 7218;
Best Local Similarity 0.9%; Pred. No. 0.0045;
Matches 3; Conservative 214; Mismatches 135; Indels 0; Gaps 0;
QY 6 ttgccagcaaatgcgaagtgagggcgccgagccgagcggtgagcccaagcccgaggg 65
Db 1440 TGTGTACRR 1381
QY 66 cgggagggcgccgctgtggaaggtctggcctggcaggaacctggcgtgggtctc 125
Db 1380 RRR 1321
QY 126 cttggccctgtgtgtgttcgcaatgctgggtgctgtgactcgtgataacctgga 185
Db 1320 RRR 1261
QY 186 gatccctgtcttggcgcaatccggggtagttgctcatcaagactagagtggtgg 245
Db 1260 RRR 1201
QY 246 aggaaggtcttcacagaaagcctgtcgtgaaatgaagagtgccaggaagcatgg 305
Db 1200 RRR 1141
QY 306 cgtgcagagaactcactcgcagaaaccacagaaacagagcagatgagga 357
Db 1140 RRR 1089
RESULT 2
US-09-165-264-8
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8
Query Match 10.6%; Score 39.2; DB 4; Length 319;
Best Local Similarity 58.6%; Pred. No. 0.15;
Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
QY 19 gcgaagtgagggcgccgagcggtgagccagccagccgagggcgagggcg 78
Db 198 ggg 257
QY 79 ggcctgtggaaggtctgtggcctggcagacctggcgtgggtctcttggccct 134
Db 258 ggg 313
RESULT 3
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155

Query Match 10.5%; Score 39; DB 3; Length 152331;
Best Local Similarity 63.2%; Pred. No. 0.62; Mismatches 60; Conservative 0; Gaps 0;
QY 27 gaggggcgggcgccgagcggtgagcccaagcccgagggcgagggcgccgtgtg 86
Db 22204 GCGGGGGGGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 22145
QY 87 ggaaggggtctggcctggcaggaacctggcgtggg 121
Db 22144 GCGGGGGGGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 22110
RESULT 4
US-09-165-264-12
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12
Query Match 10.2%; Score 37.6; DB 4; Length 318;
Best Local Similarity 56.5%; Pred. No. 0.38; Mismatches 70; Conservative 0; Gaps 0;
QY 19 gcaagtgagggcgccgagcggtgagccagccagccagggcgagggcg 78
Db 187 ggg 246
QY 79 ggcctgtggaaggtctgtggcctggcaggaacctggcgtgggtctcttggccct 138
Db 247 ggg 306
QY 139 tgtg 142
Db. 307 tgtg 310
RESULT 5
US-09-165-264-13
; Sequence 13, Application US/09165264
; Patent No. 6197510

```
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoothy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13
```

```
Query Match          10.2%; Score 37.6; DB 4; Length 320;
Best Local Similarity 59.3%; Pred. No. 0.38;
Matches 64; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 19 gcgaaggtgagggggcgccgcggtgtagccaagcccgagggcgagggcgcg 78
Db 211 gggggggggggggggggggggggggggggggggggggggggggggggggg
QY 79 ggcctgtggaaggggtctggccctggcagacctgggtggtctcc 126
Db 271 ggggggggggggggggggggggggggggggggggggggggggggggggggg
RESULT 6
US-09-165-264-14
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoothy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14
```

```
Query Match          10.1%; Score 37.2; DB 4; Length 320;
Best Local Similarity 56.6%; Pred. No. 0.48;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 19 gcgaaggtgagggggcgccgcggtgtagccaagcccgagggcgagggcgcg 78
Db 198 gggggggggggggggggggggggggggggggggggggggggggggggggg
QY 79 ggcctgtggaaggggtctggccctggcagacctgggtggtctcttggccctg 138
Db 258 gggggggggggggggggggggggggggggggggggggggggggggggggg
QY 139 tg 140
Db 318 ag 319

RESULT 7
US-09-165-264-7
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
```

```
; APPLICANT: Vinayagamoothy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

Query Match          10.0%; Score 37; DB 4; Length 320;
Best Local Similarity 47.3%; Pred. No. 0.54;
Matches 112; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 19 gcgaaggtgagggggcgccgcggtgtagccaagcccgagggcgagggcgcg 78
Db 69 gggggggggggggggggggggggggggggggggggggggggggggggggg
QY 79 ggcctgtggaaggggtctggccctggcagacctgggtggtctcttggccctg 138
Db 129 gggggggggggggggggggggggggggggggggggggggggggggggggg
QY 139 tggctttgcgcaatgctgggtgctgtgactctcgataacctggagatccctgttt 198
Db 189 ggggggggggggggggggggggggggggggggggggggggggggggggggg
QY 199 gggcgaatccggggtagtctcatcaagactagaggtgagggcgaggaaggct 255
Db 249 gggggggggggggggggggggggggggggggggggggggggggggggggg
```

```
RESULT 8
US-07-885-972A-3/c
; Sequence 3, Application US/07885972A
; Patent No. 5460950
; GENERAL INFORMATION:
; APPLICANT: Barr, Phillip J.
; APPLICANT: Brake, Anthony J.
; APPLICANT: Kaufman, Knadal J.
; APPLICANT: Tekamp-Olson, Patricia
; APPLICANT: Wasley, Louise
; APPLICANT: Wong, Polly A.
; TITLE OF INVENTION: Expression of PACE in Host Cells and
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson & Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,972A
; FILING DATE: 19920520
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/621,092
; FILING DATE: 26-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,859
; FILING DATE: 29-NOV-1990
```



```

RESULT 11
US-08-406-030A-3
; Sequence 3, Application US/08406030A
; Patent No. 6270989
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Heuge, Brian M.
; APPLICANT: Seiden, Richard F
; TITLE OF INVENTION: Protein Production and Delivery
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

RESULT 12
US-09-165-264-11
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Seq
US-09-165-264-11

```

```

Query Match          9.6%; Score 35.6; DB 4; Length 320;
Best Local Similarity 44.6%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 140; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 19 gcgaaggtgagggcgccgagggcgtagcccaagcccgagggcgagggcgagggcg 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 ggggggggggggggggggggggggggggggggggggggggggggggggggggggg 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 79 ggccttggaagggtctggccctggcagacactggctgggctctccttgccctgctg 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 139 tggcttgcgcaactgctgggtgctgactctcgataaacctggagatccctgtttt 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 199 gggcgaatccggggtagttgctcatcaagactagaggtgggggtggaaggccttca 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 259 tacagaagcctctcgcgaataaagagtggtggccagggaaagcatgcatgcagagaaac 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 319 tcactccgcagaaa 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 agatttgaagaaa 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-07-793-961A-1
; Sequence 1, Application US/07793961A
; Patent No. 5334521
; GENERAL INFORMATION:
; APPLICANT: Yoshihiro Ishikawa
; TITLE OF INVENTION: Cloning and Character-
; TITLE OF INVENTION: Iization of a Cardiac Adenylyl Cyclase
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Alan M. Gordon
; ADDRESSEE: American Cyanamid Company
; STREET: 1937 West Main Street,
; STREET: P.O. Box 60
; CITY: Stamford
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII from DW4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/793.961A
; FILING DATE: 19911118
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 31,705
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203 321 2719
; TELEFAX: 203 321 2971
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4046 base pairs listed
; TYPE: NUCLEIC ACID

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-793-961A-1

Query Match          9.6%; Score 35.4; DB 1; Length 4046;
Best Local Similarity 61.3%; Pred. No. 2.3; Indels 0; Gaps 0;
Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 27 gagggggcgggcgccgagggcgtagcccaagcccgagggcgagggcgagggcgctgtg 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 GGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 87 ggaaggggtctggccctggcaggaacctgggctgg 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GCGCGGGGTCCGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-08-240-357-1
; Sequence 1, Application US/08240357
; Patent No. 5578481
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Yoshihiro
; TITLE OF INVENTION: Cloning and Characterization of a
; TITLE OF INVENTION: Cardiac Adenylyl Cyclase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240,357
; FILING DATE: 10-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 31,705-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3244
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4046 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 131...3625
US-08-240-357-1

```

```

Query Match          9.6%; Score 35.4; DB 1; Length 4046;
Best Local Similarity 61.3%; Pred. No. 2.3; Indels 0; Gaps 0;
Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 27 gagggggcgggcgccgagggcgtagcccaagcccgagggcgagggcgagggcgctgtg 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 GGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 87 ggaaggggtctggccctggcaggaacctgggctgg 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 24, 2002, 00:24:32 ; Search time 525.6 Seconds
(without alignments)
1208.635 Million cell updates/sec

Title: US-09-696-686-51
Perfect score: 370
Sequence: 1 tgcgattgccgcaaatgc.....atgagagacgcctgcgcggcc 370

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	370	100.0	370	22	T243 gene homolog
2	42.4	11.5	4362	22	Human reproductive
3	41.8	11.3	1337	20	Human gene express
4	41.2	11.1	972	22	Human cacng8L cDNA
5	41.2	11.1	972	23	DNA encoding novel
6	40	10.8	38186	20	Human METH1 relate
7	40	10.8	38186	22	AC004449 cDNA clon
8	39.6	10.7	114955	20	Human adenosine A1
9	39	10.5	3198	20	Human IL-1ra BAC c

C 10	38.6	10.4	10828	22	AA07171	Human reproductive
C 11	38.4	10.4	357	22	AA181967	Human polynucleoti
C 12	38.2	10.3	1000	21	AAA02484	Human colon cancer
C 13	37.6	10.2	318	21	AAA38184	Primer used in the
C 14	37.6	10.2	320	21	AAA38185	Primer used in the
C 15	37.4	10.1	921	23	AA567438	DNA encoding novel
C 16	37.4	10.1	1286	23	AB199656	Mouse ischaemic co
C 17	37.4	10.1	1315	20	AAZ41312	Human normal ovari
C 18	37.4	10.1	21747	22	AAZ41312	Human immune/haema
C 19	37.2	10.1	320	22	AAK76413	Primer used in the
C 20	37.2	10.1	419	23	AAK38186	DNA encoding novel
C 21	37.2	10.1	154746	24	AA580334	Human herpesvirus
C 22	37.2	10.1	154746	24	AA580334	Human colon cancer
C 23	37	10.0	2152	22	AAH15432	Human ovarian tumo
C 24	36.8	9.9	1593	21	AAK02504	M. tuberculosis an
C 25	36.8	9.9	2188	20	AAZ77506	M. tuberculosis re
C 26	36.6	9.9	1166	20	AAZ19440	Corn Wuschel (WUS)
C 27	36.6	9.9	1166	20	AAZ19440	DNA encoding novel
C 28	36.4	9.8	513	22	AA502228	Human immune/haema
C 29	36.4	9.8	792	23	AA580335	Human gene express
C 30	36.4	9.8	3485	23	AA581530	Human polynucleoti
C 31	36.4	9.8	10468	22	AAK77961	Human polynucleoti
C 32	36.2	9.8	434	22	AAI84343	Bovine-derived cDNA
C 33	36.2	9.8	1312	20	AAZ17264	Mouse ischaemic co
C 34	36	9.7	349	22	AAI84504	Human secreted pro
C 35	36	9.7	1008	22	AAH21004	PACE composite seq
C 36	36	9.7	1045	24	AB199807	Human PACE coding
C 37	36	9.7	1529	21	AB199807	Human PACE-like pr
C 38	36	9.7	4405	13	AAQ25444	Human PACE encodin
C 39	36	9.7	4405	16	AAQ25444	cDNA encoding nove
C 40	36	9.7	4405	21	AAZ25275	Human immune/haema
C 41	36	9.7	4405	21	AAZ38795	Human METH2 relate
C 42	36	9.7	6288	22	AA545090	Human METH2 relate
C 43	36	9.7	7213	20	AAK85121	Human METH2 relate
C 44	36	9.7	8670	20	AAZ32059	Human METH2 relate
C 45	36	9.7	8670	22	AAZ32059	Human METH2 relate

ALIGNMENTS

RESULT 1
AAS05285
ID AAS05285 standard; DNA; 370 BP.
AC AAS05285;
XX AAS05285;
DT 07-SEP-2001 (first entry)
XX T243 gene homologous sequence #2 generated by PCR.
DE Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW Transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; ds.
XX Mus sp.
OS Mus sp.
PN WO200130798-A1.
XX WO200130798-A1.
PD 03-MAY-2001.
XX 03-MAY-2001.
PF 26-OCT-2000; 2000WO-US29382.
XX 26-OCT-1999; 99US-0161488.
PR (DELT-) DELTAGEN INC.
XX Klein R, Matthews W, Moore M, Allen KD;
PI WPI; 2001-300473/31.
DR Novel transgenic animals useful as animal model for characterization of
XX function of a gene encoding trinucleotide repeat proteins (TRPs),
PT


```

PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240360.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX Disclosure; SEQ ID NO 8483; 1297pp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention.
XX Sequence 4362 BP; 978 A; 1225 C; 1182 G; 977 T; 0 other;

Query Match 11.5%; Score 42.4; DB 22; Length 4362;
Best Local Similarity 58.9%; Pred. No. 0.15;
Matches 73; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Oy 11 cagcaaatcgaaagtgggggggggggggggggggggggggggggggggggggggggggggga 70
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3785 CAGCAGCGCCTGAGGGGGGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 3726

Oy 71 ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 130
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3725 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 130

Oy 131 ecct 134
    |||
Db 3665 GCCT 3662

RESULT 3
AAZ17263/c
ID AAZ17263 standard; cDNA; 1337 BP.
XX AC AAZ17263;
XX DT 12-OCT-1999 (first entry)
XX DE Human gene expression product cDNA sequence SEQ ID NO:4735.
XX KW Human; gene: gene expression product; diagnosis; therapy; probe;
XX KW detection; mapping; tissue typing; profiling; forensic; cancer;
XX XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX OS Homo sapiens.
XX XX WO9938972-A2.
XX XX 05-AUG-1999.
XX XX 28-JAN-1999; 99WO-US01619.
XX XX 03-APR-1998; 98US-0080566.
XX XX 28-JAN-1998; 98US-0072910.
XX XX 24-FEB-1998; 98US-0075954.
XX XX 31-MAR-1998; 98US-0080114.
XX XX 03-APR-1998; 98US-0080515.
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;
XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
```

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI; 1999-494092/41.
DR Novel human genes and their expression products which are
XX differentially expressed in different cell types
PT
XX
XX
PS Claim 1; Page 2250-2251; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AA212532 to AA217779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
XX Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 669 other;
SQ

Query Match 11.3%; Score 41.8; DB 20; Length 1337;
Best Local Similarity 38.7%; Pred. No. 0.16;
Matches 91; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
QY 19 qcgaagtgaaggcgccgagcgccgagcagccagccagccgagcggcgagggcgccg 78
Db 749 GGGGNGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 690
QY 79 ggcctgtggaaagggtgtggcctgcagagacctggcctggcctgtctctctgtcctg 138
Db 689 NGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 630
QY 139 tgtccttgccgcaatgctggtgctgactctcgtcgataaactcgagatccctgctttt 198
Db 629 NGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 570
QY 199 ggcgaatccggggggtgtgtctcatcaagactagaggtgggggtggaggaagg 253
Db 569 NNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 515

RESULT 4
ID AAF81304
XX AAF81304 standard; cDNA; 972 BP.
XX
XX AAF81304;
XX
XX 01-JUN-2001 (first entry)
XX
XX Human cacng8L cDNA.
XX
XX Human: starazain-like; CACNG8L; anticonvulsant; hypotensive; vasotropic;
KW antiarrhythmic; antianginal; analgesic; nootropic; cytostatic;
KW neuroprotective; cerebroprotective; antidiabetic; antimigraine;
KW antiarteriosclerotic; immunosuppressive; antiparkinsonian; auditory;
KW ophthalmological; tranquiliser; neuronal disorder;
KW calcium channel modulation; ss.
XX Homo sapiens.
OS

XX WO200121791-A2.
XX
XX 29-MAR-2001.
XX
XX 25-SEP-2000; 2000WO-GB03685.
XX
XX 23-SEP-1999; 99GB-0022571.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX Clare JJ, Plumpton M, Moss FJ, Sanseau P;
XX WPI; 2001-266074/27.
XX P-PSDB; AAB73982.
XX
XX Novel starazain-like polypeptides capable of modulating the steady
PT state inactivation of an alpha pore-forming subunit of a voltage-gated
PT calcium channel, for treating central nervous system disorders -
XX
XX Claim 7; Page 45-47; 48pp; English.
XX
XX The present sequence is provided in a specification relating to an
CC isolated starazain-like polypeptide comprising a sequence of 327 or 275
CC amino acids fully defined in the specification, or its variant capable
CC of modulating the steady state. The polypeptide is useful for treating
CC a subject having a neuronal disorder that is responsive to calcium
CC channel modulation. It is useful in the manufacture of a medicament for
CC treatment or prophylaxis of a disorder that is responsive to calcium
CC channel modulation, e.g. epilepsy, episodic ataxia, spinocerebellar
CC ataxia, hypertension, ischaemic heart disease, arrhythmia, angina,
CC pain, cerebral ischaemia, Alzheimer's disease, neuroprotection, stroke,
CC diabetes, cerebral vasospasm, atherosclerosis, tardive dyskinesias,
CC peripheral vascular disease, immunosuppression, cancerous diseases,
CC migraine, headache, bipolar disorder, unipolar depression, anxiety,
CC Parkinson's disease, cognitive disorders, ophthalmic diseases,
CC neuromuscular disorders and tinnitus. Nucleic acids encoding these
CC polypeptides are useful in the production of the polypeptides, and as
CC primers.
XX
XX Sequence 972 BP; 193 A; 291 C; 266 G; 222 T; 0 other;
SQ

Query Match 11.1%; Score 41.2; DB 22; Length 972;
Best Local Similarity 49.1%; Pred. No. 0.22;
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 3 cgattgccagcaaatgcgaagtgtgagggcgggcgggcgagcagccagccgag 62
Db 45 cgcaagccttgaaaaaggccttgaggtagtgggcaagcccaagcagtcgctggt 104
QY 63 gggcgaggagggcgggcgccctgtggaagggtctggcctgcagagacctggcgggt 122
Db 105 ggagagcgtgcgactagtgtgcacagcaacggtccaggaagatgagtgctcgaggagaa 164
QY 123 ctctctggcctgctgtgtgttgcgcaatgctgggtgctgactcgtcgataacct 182
Db 165 ggcctgacctgctgtagcagtgcttctgtgctggtggcctcctcctggtatcgc 224
QY 183 ggagatccctgcttttggcgcaatccgggggtgagtgctcat 224
Db 225 ggtcagcacccactactggtgtacctgtacctggagagggtgtgat 266

RESULT 5
ID AAS80100
XX AAS80100 standard; cDNA; 972 BP.
XX
XX AAS80100;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #15904.
DE

XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG15913.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity
XX	
PS	Claim 1; SEQ ID NO 15904; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences .
XX	
SQ	Sequence 972 BP; 193 A; 291 C; 266 G; 222 T; 0 other;

RESULT 6
AAZ32028
ID AAZ32028 standard; DNA; 38186 BP.
XX AC
AC AAZ32028;
XX DT
DT 10-JAN-2000 (first entry)
XX DE
DE Human METH1 related EST AC004449.
XX KW
KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency; ss.
XX OS
OS Homo sapiens.
XX PN
PN W09937660-A1.
XX PD
PD 29-JUL-1999.
XX PF
PF 22-JAN-1999; 99WO-US01313.
XX PR
PR 23-JAN-1998; 98US-0072298.
XX PR
PR 28-AUG-1998; 98US-0098539.
XX PA
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
XX PI
PI IrueLa-Arispe L, Hastings GA, Ruben SM;
XX WPTI; 1999-590684/50.
XX DR
DR
XX PT
PT New isolated metalloprotease thrombospondin polypeptides, useful for
PT treating hyperproliferative disorders, cancers or autoimmune disorders
XX Disclosure; Page 363-387; 457pp; English.
XX CC
CC AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC respectively. METH1 and METH2 have been found to be potent inhibitors of
CC angiogenesis both in vitro and in vivo. They can be used for treating
CC cancer and other disorders related to angiogenesis including abnormal
CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC endometrial bleeding disorders, diabetic retinopathy, some forms of
CC macula degeneration, haemangiomas, and arterial-venous malformations.
CC They may be useful in treating deficiencies or disorders of the immune
CC system, by activating or inhibiting the proliferation, differentiation,
CC or mobilisation (chemotaxis) of immune cells. The etiology of these
CC immune deficiencies or disorders may be genetic, somatic, such as
CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC toxins), or infectious. They can also be used to treat inflammatory
CC conditions, both chronic and acute conditions. The products can also be
CC used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
CC AAY49511 represent sequences given in the exemplification of the present
XX invention.

Sequence 38186 BP: 7571 A; 11503 C; 12193 G; 6919 T; 0 other;

QY 68 gtaggggagggcctgtggagaggtctggcctggcagacctgggctgggtctctct 127
 DB 2759 gtaggggagggcctgtggagaggtctggcctggcagacctgggctgggtctctct 2818
 QY 128 tggcctgtgtgtgc 143
 DB 2819 acgtggcccaaggtgc 2834

RESULT 7
 AAC90085
 ID AAC90085 standard; DNA; 38186 BP.
 XX
 AC AAC90085;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE AC004449 cDNA clone.
 XX
 KW METH; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
 KW plaque neovascularisation; telangiectasia; haemophilic joint; ESI;
 KW angiofibroma; fibromuscular dysplasia; expressed sequence tag;
 KW Crohn's disease; atherosclerosis; birth control; ss.
 XX
 OS Unidentified.

XX
 PN WO200071577-A1.
 XX
 PD 30-NOV-2000.
 XX
 PF 25-MAY-2000; 2000WO-US14462.
 XX
 PR 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONA/) JONAK Z L.
 PA (TRUL/) TRULLI S H.
 PA (FORN/) FORNWALD J A.
 PA (TERR/) TERRETT J A.
 XX
 PI IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornwald JA, Terrett JA;
 XX
 DR WPI; 2001-025136/03.
 XX
 PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 XX
 XX Claim 7; Pages 663-687; 768pp; English.
 XX
 CC The present invention relates to human METH1 and METH2, (ME for
 CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
 CC The present sequence is an expressed sequence tag (EST) for METH. METH
 CC can be used for inhibiting angiogenesis in an individual, and for
 CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC

CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joint, angiofibroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer.
 XX
 SQ Sequence 38186 BP; 7571 A; 11502 C; 12194 G; 6919 T; 0 other;

Query Match 10.8%; Score 40; DB 22; Length 38186;
 Best Local Similarity 55.9%; Pred. No. 1.1;
 Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 8 gccacagaaatgcgaaggtgagggcgccggcgccgtagcgaagggggcg 67
 DB 2699 ggcggggatgggaggggagggcgccggcgccgtagcgaagggggcg 2758
 QY 68 gtaggggagggcgccgtgggaaggtctggcctggcagacctgggctctct 127
 DB 2759 gtaggggagggcgccgtgggaaggtctggcctggcagacctgggctctct 2818
 QY 128 tggcctgtgtgtgc 143
 DB 2819 acgtggcccaaggtgc 2834

RESULT 8
 AAX53491
 ID AAX53491 standard; DNA; 114955 BP.
 XX
 AC AAX53491;
 XX
 DT 05-JUL-1999 (first entry)
 XX
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.
 XX
 KW Antisense oligonucleotide; multiple target; antisense treatment;
 KW impaired respiration; inflammation; lung disease;
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 XX
 OS Synthetic.
 XX
 PN WO9913886-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 17-SEP-1998; 98WO-US19419.
 XX
 PR 09-JUN-1998; 98US-0093972.
 PR 17-SEP-1997; 97US-0059160.
 XX
 XX (UYEC-) UNIV EAST CAROLINA.
 PA
 XX Nyce JW;
 XX
 DR WPI; 1999-229400/19.
 XX
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 XX
 PS Disclosure; Page 37; 120pp; English.
 XX

CC The specification describes antisense oligonucleotides (AA52869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences AAX5272-74. These multiple target oligonucleotides (specifically AAX5180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, CC hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer.

XX
SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 10.7%; Score 39.6; DB 20; Length 114955;
Best Local Similarity 39.4%; Pred. No. 1.7;
Matches 97; Conservative 15; Mismatches 134; Indels 0; Gaps 0;

QY 2 gcgattgccagcaaatgcgaagtgtagggcgccggcgccgctgacgaagccga 61
DB 108020 gcgctcgctcctbgbggcccsnnndnngcgggcgccgctgctgcctbg 108079
QY 62 gggcgcgaggggcgccgctgtgggaaggtctctggcctgcaggaacctggctggg 121
DB 108080 gggccsnnndnngcgggcgccgctgctgcctbgbggccsnnndnngcgggg 108139
QY 122 tctcttgccctgtgtgtgtgttgcggaatgtcgtgtgtgtcgtcgtgataacc 181
DB 108140 gcggcgccctggctgcctbgbggccsnnndnngcgggcgccgctgctgcgc 108199
QY 182 tggagatccctgtcttttggcgcaatccgggggtagtgtctcatcaagactaggtggg 241
DB 108200 tbgggennndnngcgggcgccgctgctgcctbgbggccsnnndnngcgggg 108259
QY 242 gtggag 247
DB 108260 gcggcg 108265

RESULT 9
AAX02974/c
ID AAX02974 standard; DNA; 3198 BP.
XX
AC AAX02974;
XX
DT 22-JUN-1999 (first entry)
XX
DE Human IL-1ra BAC contiguous DNA sequence 19.
XX
KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS Homo sapiens.
XX
PN WO9906426-A1.
XX
PD 11-FEB-1999.
XX
PF 03-AUG-1998; 98WO-US16102.
XX
PR 02-JUL-1998; 98US-0091650.

PR 04-AUG-1997; 97US-0054646.
XX
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI Pan Y;
XX
XX WPI; 1999-153692/13.
DR
PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT used to inhibit inflammation and to screen for specific modulators
XX
PS Example 5; Figure 3; 226pp; English.
XX
CC AAX02956-X03048 and AAX22301-X22304 are overlapping BAC genomic
CC sequences containing alternatively spliced forms of human IL-1ra. Such
CC fragments are used in the method of the invention which describes the
CC isolation of a novel human TANGO-77 encoding nucleic acid and protein.
CC Tango-77 is a member of the cytokine superfamily that is expected to
CC inhibit inflammation by binding to the interleukin-1 receptor (IL-1R). It
CC may also bind to a new receptor so could regulate other cellular
CC processes associated with acute or chronic inflammation, e.g. asthma,
CC chronic myelogenous leukaemia, rheumatoid arthritis, psoriasis and
CC inflammatory bowel disease. It may also induce or suppress interleukins,
CC cytokines and growth factors. Modulators of this protein are used to
CC treat or prevent conditions associated with abnormal levels of
CC inflammation, or activity of IL-1 or its receptor complex.
XX
SQ Sequence 3198 BP; 807 A; 847 C; 681 G; 846 T; 17 other;

Query Match 10.5%; Score 39; DB 20; Length 3198;
Best Local Similarity 63.2%; Pred. No. 1.1;
Matches 60; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 27 gagggggcgggcgccgctgtagcccaagccgagggcgggcgggcgggcgctgtg 86
DB 3113 GCGGGGGGGCGCGGG 3054
QY 87 ggaagggctgtggcctggcaggacctggctgggg 121
DB 3053 GGGGGGGGGCGGG 3019

RESULT 10
AAL07171/c
ID AAL07171 standard; DNA; 10828 BP.
XX
AC AAL07171;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9859.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205315.
PR 07-JUN-2000; 2000US-0209467.

PR	28-JUN-2000;	2000US-0214886.	PR	20-OCT-2000;	2000US-0241786.
PR	30-JUN-2000;	2000US-0215135.	PR	20-OCT-2000;	2000US-0241787.
PR	07-JUL-2000;	2000US-0216647.	PR	20-OCT-2000;	2000US-0241808.
PR	07-JUL-2000;	2000US-0216880.	PR	20-OCT-2000;	2000US-0241809.
PR	11-JUL-2000;	2000US-0217487.	PR	20-OCT-2000;	2000US-0241826.
PR	11-JUL-2000;	2000US-0217496.	PR	01-NOV-2000;	2000US-0244617.
PR	11-JUL-2000;	2000US-0218290.	PR	08-NOV-2000;	2000US-0246474.
PR	26-JUL-2000;	2000US-0220963.	PR	08-NOV-2000;	2000US-0246475.
PR	26-JUL-2000;	2000US-0220964.	PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0224519.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0225213.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225214.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0225266.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000;	2000US-0225267.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0246610.
PR	18-AUG-2000;	2000US-0226279.	PR	08-NOV-2000;	2000US-0246611.
PR	22-AUG-2000;	2000US-0226681.	PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000;	2000US-0226868.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249208.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249209.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249210.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249214.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249216.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249218.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232081.	PR	17-NOV-2000;	2000US-0249299.
PR	12-SEP-2000;	2000US-0232196.	PR	17-NOV-2000;	2000US-0249300.
PR	14-SEP-2000;	2000US-0232397.	PR	01-DEC-2000;	2000US-0250160.
PR	14-SEP-2000;	2000US-0232398.	PR	01-DEC-2000;	2000US-0250160.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232400.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232401.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0233063.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251868.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251869.
PR	21-SEP-2000;	2000US-0234274.	PR	08-DEC-2000;	2000US-0251989.
PR	25-SEP-2000;	2000US-0234397.	PR	08-DEC-2000;	2000US-0251990.
PR	25-SEP-2000;	2000US-0234998.	PR	11-DEC-2000;	2000US-0254097.
PR	26-SEP-2000;	2000US-0234998.	PR	05-JAN-2001;	2001US-0259678.
PR	26-SEP-2000;	2000US-0235484.	XX		
PR	27-SEP-2000;	2000US-0235836.	XX		
PR	27-SEP-2000;	2000US-0235836.	XX		
PR	29-SEP-2000;	2000US-0236327.	XX		
PR	29-SEP-2000;	2000US-0236327.	XX		
PR	29-SEP-2000;	2000US-0236367.	XX		
PR	29-SEP-2000;	2000US-0236368.	XX		
PR	29-SEP-2000;	2000US-0236369.	XX		
PR	29-SEP-2000;	2000US-0236370.	XX		
PR	29-SEP-2000;	2000US-0236370.	XX		
PR	02-OCT-2000;	2000US-0236802.	XX		
PR	02-OCT-2000;	2000US-0237037.	XX		
PR	02-OCT-2000;	2000US-0237038.	XX		
PR	02-OCT-2000;	2000US-0237039.	XX		
PR	02-OCT-2000;	2000US-0237040.	XX		
PR	13-OCT-2000;	2000US-0239935.	XX		
PR	13-OCT-2000;	2000US-0239937.	XX		
PR	20-OCT-2000;	2000US-0240960.	XX		
PR	20-OCT-2000;	2000US-0241221.	XX		
PR	20-OCT-2000;	2000US-0241785.	XX		

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen - is used in preventing, treating or ameliorating a medical condition -

Disclosure; SEQ ID NO 9859; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention.

Sequence 10828 BP; 2518 A; 3387 C; 2921 G; 2002 T; 0 other;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:18:14 ; Search time 3053.4 Seconds
(without alignments)
1958.521 Million cell updates/sec

Title: US-09-696-686-51
Perfect score: 370
Sequence: 1 tgcgattgccagcaaatgc.....atgaggagccctgccgccc 370

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	GenEmbl :
1: gb_ba.*	AC074220 Mus muscu
2: gb_htg.*	AC102804 Homo sapi
3: gb_in.*	AC106591 Rattus no
4: gb_om.*	I66494 Sequence 14
5: gb_ov.*	U05204 Oryctolagus
6: gb_pat.*	AC023197 Mus muscu
7: gb_ph.*	AC094865 Rattus no
8: gb_pl.*	AC095701 Rattus no
9: gb_pr.*	AC095225 Rattus no
10: gb_ro.*	AC100281 Mus muscu
11: gb_sts.*	AC093968 Rattus no
12: gb_sy.*	AC095175 Rattus no
13: gb_un.*	AC079431 Mus muscu
14: gb_vl.*	AC094597 Rattus no
15: em_ba.*	AC084799 Mus muscu
16: em_fun.*	AC100680 Mus muscu
17: em_hum.*	AC100569 Mus muscu
18: em_in.*	AC101154 Mus muscu
19: em_mu.*	AC022588 Homo sapi
20: em_om.*	AC021929 Homo sapi
21: em_or.*	AC016280 Homo sapi
22: em_ov.*	AC090865 Homo sapi
23: em_pat.*	AC097363 Rattus no
24: em_ph.*	Continuation (2 of
25: em_pl.*	AC163301 Homo sapi
26: em_pr.*	AC022595 Homo sapi
27: em_ro.*	AC095300 Rattus no
28: em_sts.*	AC027682 Homo sapi
29: em_sy.*	AC087563 Homo sapi
30: em_un.*	AL395019 T3 end of
31: em_vl.*	AC027483 Homo sapi
32: em_mu.*	AC107595 Rattus no
33: em_om.*	AC106169 Rattus no
34: em_or.*	AC087337 Homo sapi
35: em_ov.*	AC101222 Mus muscu
36: em_pat.*	AC018939 Homo sapi
37: em_ph.*	AC028508 Homo sapi
38: em_pl.*	AC022783 Mus muscu
39: em_pr.*	AC079640 Mus muscu
40: em_ro.*	AC107846 Mus muscu
41: em_sts.*	AC105985 Mus muscu
42: em_sy.*	AC021584 Homo sapi
43: em_un.*	AC074204 Mus muscu
44: em_vl.*	AC007610 Homo sapi
45: em_mu.*	AC079429 Mus muscu

ALIGNMENTS

RESULT 1
AC074220/c

LOCUS
DEFINITION
AC074220
AC074220.7 GI:13876390
HTG; HTGS_PHASE1; HTGS_DRAFT.
house mouse.
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.
AC074220
AC074220.7 GI:13876390
HTG; HTGS_PHASE1; HTGS_DRAFT.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 89855)
Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 89855)
Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
-----	-----	-----	-----	-----

Fri May 24 10:49:50 2002

JOURNAL

Submitted (20-JUL-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Apr 28, 2001 this sequence version replaced qi:11276104.

On Apr 28. 2001 this sequence replaced qi:11276104.

COMMENT On Apr 28, 2001 this sequence was
-----Genome Center

-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC

Center Code: HPGC

Web site:
<http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mouse.html>

e.html
contact: anktm@cs.necad.bwh.harvard.edu

Contact: gnktm@capecod.bwh.harvard.
-----Summary Statistics

-----Summary-----
Center project name: ADY

Center project name: ADI
sequencing vector: pUC18; L08752

sequencing vector: pOC16, Eos/32
Chemistry: Dye-terminator Big Dye; 100%

***Consensus quality:** 79355 at least Q20

*Consensus quality:	75355 at least Q20
*Consensus quality:	75365 at least Q30

*Consensus quality: 69022 at least Q40

Estimated insert size: agarose-FP - N/

****Estimated insert size: 89115 - sum-of-contigs**

Quality coverage: agarose-FP - N/A

Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation

Quality coverage:
-----,

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

7963:	contig of 7962 bp in length
7982:	gap of unknown length
7983:	contig of 5942 bp in length
13925:	gap of unknown length
13945:	contig of 5046 bp in length
18991:	gap of unknown length
19011:	contig of 3491 bp in length
22501:	contig of 3491 bp in length
22502:	contig of unknown length
22521:	gap of unknown length
27077:	contig of 4556 bp in length
27097:	gap of unknown length
28819:	contig of 1722 bp in length
28839:	gap of unknown length
28820:	gap of unknown length
31224:	contig of 2385 bp in length
31225:	gap of unknown length
31244:	gap of unknown length
35093:	contig of 3849 bp in length
35113:	gap of unknown length
35114:	contig of 3498 bp in length
38612:	gap of unknown length
38632:	contig of 3630 bp in length
42262:	gap of unknown length
42281:	gap of unknown length
44130:	contig of 1849 bp in length
44131:	gap of unknown length
44150:	gap of unknown length
47059:	contig of 2909 bp in length
47079:	gap of unknown length
47080:	contig of 2417 bp in length
49496:	gap of unknown length
49497:	gap of unknown length
52336:	contig of 2720 bp in length
52337:	gap of unknown length
52356:	gap of unknown length
54270:	contig of 2014 bp in length
54271:	gap of unknown length
54290:	gap of unknown length
56271:	contig of 1981 bp in length
56272:	gap of unknown length
56291:	gap of unknown length
58849:	contig of 2558 bp in length
58850:	gap of unknown length
58870:	contig of 1752 bp in length
60622:	gap of unknown length
60641:	gap of unknown length
62255:	contig of 1614 bp in length
62275:	gap of unknown length
63085:	contig of 810 bp in length
63105:	gap of unknown length
63106:	contig of 2335 bp in length
65440:	gap of unknown length
65441:	gap of unknown length

*	65461	67629: contig of 2169 bp in length
*	67630	67649: gap of unknown length
*	67650	69178: contig of 1529 bp in length
*	69179	69198: gap of unknown length
*	69199	70095: contig of 897 bp in length
*	70096	70115: gap of unknown length
*	70116	71821: contig of 1706 bp in length
*	71822	71841: gap of unknown length
*	71842	73506: contig of 1665 bp in length
*	73507	73526: gap of unknown length
*	73527	74749: contig of 1223 bp in length
*	74750	74769: gap of unknown length
*	74770	75963: contig of 1194 bp in length
*	75964	75983: gap of unknown length
*	75984	77905: contig of 1922 bp in length
*	77906	77925: gap of unknown length
*	77926	79236: contig of 1311 bp in length
*	79237	79256: gap of unknown length
*	81332	81332: contig of 2076 bp in length
*	81333	81352: gap of unknown length
*	81353	83169: contig of 1817 bp in length
*	83170	83189: gap of unknown length
*	83190	84096: contig of 907 bp in length
*	84117	84116: gap of unknown length
*	85750	85749: contig of 1633 bp in length
*	85770	85769: gap of unknown length
*	86771	86770: contig of 1001 bp in length
*	86791	86770: gap of unknown length
*	88208	88207: contig of 1417 bp in length
*	88228	88227: gap of unknown length
*	88977	88976: contig of 749 bp in length
*	88997	88996: gap of unknown length
*	88997	88955: contig of 859 bp in length.
FEATURES		
source		
1. .89855		
/organism="Mus musculus"		
/strain="C57BL6/J"		
/db_xref="taxon:10090"		
/chromosome="17"		
/clone="RP23-76116"		
/sex="male"		
1. .7962		
misc_feature	/note="assembly_name:Contig145"	
misc_feature	7983. .13924	
misc_feature	/note="assembly_name:Contig144"	
misc_feature	13945. .18990	
misc_feature	/note="assembly_name:Contig143"	
misc_feature	19011. .22501	
misc_feature	/note="assembly_name:Contig142"	
misc_feature	22522. .27077	
misc_feature	/note="assembly_name:Contig141"	
misc_feature	27098. .28819	
misc_feature	/note="assembly_name:Contig140"	
misc_feature	28840. .31224	
misc_feature	/note="assembly_name:Contig139"	
misc_feature	31245. .35093	
misc_feature	/note="assembly_name:Contig138"	
misc_feature	35114. .38611	
misc_feature	/note="assembly_name:Contig137"	
misc_feature	38632. .42261	
misc_feature	/note="assembly_name:Contig136"	
misc_feature	42282. .44130	
misc_feature	/note="assembly_name:Contig135"	
misc_feature	44151. .47059	
misc_feature	/note="assembly_name:Contig134"	
misc_feature	47080. .49496	
misc_feature	/note="assembly_name:Contig133"	
misc_feature	49517. .52236	
misc_feature	/note="assembly_name:Contig132"	
misc_feature	52257. .54270	
misc_feature	/note="assembly_name:Contig131"	
misc_feature	54291. .56271	
misc_feature	/note="assembly_name:Contig130"	


```

misc_feature 56292..58849
              /note="assembly_name:Contig129"
misc_feature 58870..60621
              /note="assembly_name:Contig128"
misc_feature 60642..62255
              /note="assembly_name:Contig127"
misc_feature 62276..63085
              /note="assembly_name:Contig126"
misc_feature 63106..65440
              /note="assembly_name:Contig125"
misc_feature 65461..67629
              /note="assembly_name:Contig124"
misc_feature 67650..69178
              /note="assembly_name:Contig123"
misc_feature 69199..70095
              /note="assembly_name:Contig122"
misc_feature 70116..71821
              /note="assembly_name:Contig121"
misc_feature 71842..73506
              /note="assembly_name:Contig120"
misc_feature 73527..74749
              /note="assembly_name:Contig119"
misc_feature 74770..75963
              /note="assembly_name:Contig118"
misc_feature 75984..77905
              /note="assembly_name:Contig117"
misc_feature 77926..79236
              /note="assembly_name:Contig116"
misc_feature 79257..81332
              /note="assembly_name:Contig115"
misc_feature 81353..83169
              /note="assembly_name:Contig114"
misc_feature 83190..84096

```

Query Match 99.6%; Score 368.4; DB 2; Length 89855;
 Best Local Similarity 99.7%; Pred. No. 2.5e-73;
 Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 tgcgattgccagcaaatgcgaaggtagggggcgccggggcgtagccaaagcccg 60
Db 5896 TGCATTTGCCAGCAATGCGAAGGTGAGGGCGGGCGCGGCGTAGCCAGCCCG 5837

QY 61 agggggcgaggggggcgccctgtgggaaggtctggccctgcgaagcctggcctgg 120
Db 5836 AGGGCGGGAGGGGGCGCCCTGTGCGAAGGTCTGGGCGCTGGCAGGACCTGGCTGG 5777

QY 121 gtctcttggccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
Db 5776 GTCTCTTGGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5717

QY 181 ctggagatccctgtcttttggcgaatccgggggttagtgcatacaagactagagtg 240
Db 5716 CTGGAGATCCCTGTCTTTTGGCGAATCCGGGGGTAGTTGCTCATCAAGACTAGAGTGG 5657

QY 241 ggtggagggaagcttcatacaggaagcctgtcgcgaatgaagattggcaggaagaa 300
Db 5656 GGTGGAGGGAAGCCTTCATACAGGAAGCCTGCTGCGAATGAAGAGTTGGCCAGGGAAG 5597

QY 301 catggcgtcagaggaactcactccgcagaaaccacagaaacagagcagatgagga 360
Db 5596 CATGGCGTCCAGAGAACTCACTCCGCAGAAACACAGAAACAGAGCAGATGAGGAGGC 5537

QY 361 cctgcgcggcc 370
Db 5536 CTGCGCGGCC 5527

```

RESULT 2
 AC102804/c
 LOCUS
 DEFINITION Homo sapiens chromosome 18 clone RP11-605C5 map 18, LOW-PASS
 SEQUENCE SAMPLING.
 ACCESSION AC102804

VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AC102804.1 GI:17060914
 HTG: HTGS_PHASE0.
 human

Homo sapiens

Eukaryota; Metazoa; Chordata; Creniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 53178)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 18, clone RP11-605C5

Unpublished

2 (bases 1 to 53178)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferrelira,P., Fitzflugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Glend,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidorn,J.,
 Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliou,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://fcp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L19709

Center clone name: 605_C_5

* NOTE: This record contains 65 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

```

1 739: contig of 739 bp in length
740 839: gap of 100 bp
840 1548: contig of 709 bp in length
1549 1648: gap of 100 bp in length
1649 2365: contig of 717 bp in length
2366 2465: gap of 100 bp
2466 3173: contig of 708 bp in length
3174 3273: gap of 100 bp
3274 4026: contig of 753 bp in length
4027 4126: gap of 100 bp in length
4127 4882: contig of 756 bp in length
4883 4982: gap of 100 bp
4983 5601: contig of 619 bp in length
5602 5701: gap of 100 bp

```



```

AC106591/c
LOCUS
DEFINITION
AC106591
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 39811)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alzbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbieri,J.,
Benton,J., Blimie,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brivega,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Cartier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Deigado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homai,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., McInerney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,G., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Slason,I., Sodergren,E., Sonaikar,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlaczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 39811)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Center project name: GLLB
Center clone name: CH230-212K18
-----
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 38345 bases at least Q40
Consensus quality: 42938 bases at least Q30
-----
Consensus quality: 46037 bases at least Q20
Estimated insert size: 24639; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.3x in Q20 bases; sum-of-contigs estimation
-----
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 25 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
3585: contig of 3585 bp in length
3685: gap of unknown length
3686: contig of 1612 bp in length
5297: gap of unknown length
5298: contig of 2129 bp in length
7526: gap of unknown length
7527: contig of 1690 bp in length
9316: gap of unknown length
9317: contig of 1831 bp in length
9417: gap of unknown length
11248: contig of 2193 bp in length
11348: gap of unknown length
13541: contig of 1439 bp in length
13640: gap of unknown length
15079: contig of 1490 bp in length
15179: gap of unknown length
15668: contig of 1354 bp in length
16763: gap of unknown length
16770: contig of 1484 bp in length
18123: gap of unknown length
18224: contig of 1332 bp in length
19708: gap of unknown length
19807: contig of 1179 bp in length
21139: gap of unknown length
21240: contig of 1513 bp in length
22419: gap of unknown length
22519: contig of 1145 bp in length
24032: gap of unknown length
25277: contig of 1065 bp in length
25376: gap of unknown length
26441: gap of unknown length
26442: contig of 1633 bp in length
28342: gap of unknown length
28375: contig of 1094 bp in length
29369: gap of unknown length
30475: contig of 1007 bp in length
30575: gap of unknown length
32062: contig of 1487 bp in length
32162: gap of unknown length
33609: contig of 1447 bp in length
33610: gap of unknown length
33710: contig of 1538 bp in length
35347: gap of unknown length
35348: contig of 1019 bp in length
36366: gap of unknown length
36466: contig of 1011 bp in length
37477: gap of unknown length
37478: contig of 1032 bp in length
38510: gap of unknown length
38710: contig of 1102 bp in length.
Location/Qualifiers
1..39811
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-212K18"
10663 a 8757 c 7160 g 10537 t 2694 others
BASE COUNT
ORIGIN
FEATURES
Source

```

Query Match	13.78;	Score 50.6;	DB 2;	Length 39811;
Best Local Similarity	58.2;	Pred. No. 0.094;		
Matches	89;	Conservative	0;	Mismatches 64; Indels 0; Gaps 0;
QY	19	gcgaaggtgagggcgccggcgccgtagccaagccgagggcgagggcgcg 78		
Db	36123	GGGTGGGTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36064		
QY	79	ggcctgtggaaaggtctggccctggcaagacctggctgggtctctctgtgacctgtg 138		
Db	36063	GGCCGCGCGGGCGCGGGGGCGCTGGCGCGGGGGGTGGTGGGGGGCGCGCGGGGGCGCTG 36004		
QY	139	tgtgcttgcgcaatgctgggtgctgtgactc 171		
Db	36003	GGTGTGCGGCGCGCGCGCGCGCGCGGTGCGGGCTC 35971		
RESULT	4			
LOCUS	I66494	7218 bp	DNA	linear
DEFINITION	Sequence 14 from patent US 5670367.			PAT 28-DEC-1997
ACCESSION	I66494			
VERSION	I66494.1	GI:2724471		
KEYWORDS	Unknown.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	1 (bases 1 to 7218)			
AUTHORS	Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.			
TITLE	Recombinant fowlpox virus			
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;			
FEATURES	Location/Qualifiers			
Source	1..7218			
BASE COUNT	1444 a 1491 c 1486 g 1929 t	368 others		
ORIGIN				
Query Match	13.6%;	Score 50.4;	DB 6;	Length 7218;
Best Local Similarity	0.9%;	Pred. No. 0.13;		
Matches	3;	Conservative	214;	Mismatches 135; Indels 0; Gaps 0;
QY	6	ttgccagcaaaatcgaaagtgaaggcgccggcgcccgtagccagcccgagggg 65		
Db	1440	TGTTACRR 1381		
QY	66	cggagggcgccggcctgtggaaaggctgtggcctgcgagacctgggctgggtctc 125		
Db	1380	RR 1321		
QY	126	cttggccctgctgtgctttgcgcaatgctgggtgctgactctcgataacctgga 185		
Db	1320	RR 1261		
QY	186	gatccctgtttggcgaaatccgggggtagtgtcctcatcaagactagagtggggtg 245		
Db	1260	RR 1201		
QY	246	aggaaggtctcatcaggaagcctgtcgaaatagaagattggccagggaaacatgg 305		
Db	1200	RR 1141		
QY	306	gtgcagagggaactcactccgcagaaaccagaaacagagcgatgagga 357		
Db	1140	RR 1089		
RESULT	5			
LOCUS	OCU05204/c	1396 bp	mRNA	linear
DEFINITION	Oryctolagus cuniculus preproacrosin mRNA, complete cds.			MAM 27-JAN-1995
ACCESSION	U05204			
VERSION	U05204.1	GI:451841		

Db 969 GTGCCAGAGCGGCTGGTACCGCGGGGATCGGTAGTGGAGGGTAGGGGTAGC 910
Qy 259 tacaggaagcctgctgcgaatgaagagtggtgcccaggaagacatggtgctgacg 312
Db 909 TGGTTGAATCATCGCCAAAGGCATTAGAGCCGATTCTAGAACGCGATCCAGTCCAG 856

RESULT 6
AC023197/c
LOCUS AC023197 141892 bp DNA linear HTG 16-MAR-2000
DEFINITION Mus musculus clone RP22-239F2, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC023197
VERSION AC023197.2 GI:7249227
KEYWORDS HTG; HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bida, F.,
TITLE Boguslavskiy, L., Boukhgaiter, B., Brown, A., Burkett, G., Castle, A.,
JOURNAL Choepe, Y., Collangelo, M., Collins, S., Collimore, A., Cooke, P.,
REFERENCE DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
AUTHORS Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A., and Zody, M.

Direct Submission
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6957784.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5917
Center clone name: 239_F_2

* NOTE: This record contains 155 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 657 756: contig of 656 bp in length
* 757 1510: contig of 754 bp in length
* 1511 1610: gap of 100 bp
* 1611 2364: contig of 754 bp in length

2365 2464: gap of 100 bp
2465 3234: contig of 770 bp in length
3235 3334: gap of 100 bp
3335 4083: contig of 749 bp in length
4084 4183: gap of 100 bp
4184 4931: contig of 748 bp in length
4932 5031: gap of 100 bp
5032 5787: contig of 756 bp in length
5788 5887: gap of 100 bp
5888 6644: contig of 757 bp in length
6645 6744: gap of 100 bp
6745 7516: contig of 772 bp in length
7517 7616: gap of 100 bp
7617 8376: contig of 760 bp in length
8377 8476: gap of 100 bp
8477 9227: contig of 751 bp in length
9228 9327: gap of 100 bp
9328 10089: contig of 762 bp in length
10090 10189: gap of 100 bp
10190 10952: contig of 763 bp in length
10953 11052: gap of 100 bp
11053 11826: contig of 774 bp in length
11827 11925: gap of 100 bp
11927 12688: contig of 762 bp in length
12689 12788: gap of 100 bp
12789 13543: contig of 755 bp in length
13544 13643: gap of 100 bp
13644 14407: contig of 764 bp in length
14408 14507: gap of 100 bp
14508 15275: contig of 768 bp in length
15276 15375: gap of 100 bp
15376 16144: contig of 769 bp in length
16145 16244: gap of 100 bp
16245 17002: contig of 758 bp in length
17003 17102: gap of 100 bp
17103 17803: contig of 701 bp in length
17804 17903: gap of 100 bp
17904 18618: contig of 715 bp in length
18619 18718: gap of 100 bp
18719 19486: contig of 768 bp in length
19487 19586: gap of 100 bp
19587 20356: contig of 770 bp in length
20357 20456: gap of 100 bp
20457 21222: contig of 766 bp in length
21223 21322: gap of 100 bp
21323 22069: contig of 747 bp in length
22070 22169: gap of 100 bp
22170 22921: contig of 752 bp in length
22922 23021: gap of 100 bp
23022 23772: contig of 751 bp in length
23773 23872: gap of 100 bp
23873 24634: contig of 762 bp in length
24635 24734: gap of 100 bp
24735 25363: contig of 629 bp in length
25364 25463: gap of 100 bp
25464 26227: contig of 764 bp in length
26228 26327: gap of 100 bp
26328 27088: contig of 761 bp in length
27089 27188: gap of 100 bp
27189 27955: contig of 767 bp in length
27956 28055: gap of 100 bp
28056 28836: contig of 781 bp in length
28837 28936: gap of 100 bp
28937 29702: contig of 766 bp in length
29703 29802: gap of 100 bp
29803 30577: contig of 775 bp in length
30578 30677: gap of 100 bp
30678 31482: contig of 805 bp in length
31483 31583: gap of 100 bp
31583 32341: contig of 759 bp in length
32342 32441: gap of 100 bp
32442 33202: contig of 761 bp in length
33203 33302: gap of 100 bp

Query Match	13.2%	Score 48.8;	DB 2;	Length 236301;	
Best Local Similarity	50.4%;	Pred. No. 0.2;	Mismatches 117;	Indels 0;	Gaps 0;
Matches 119;	Conservative				
QY	18	tgcgaagtgagggcgggcgccgcgggcgtagccaagcccgagggcgggagggggcg 77			
Db	89602	TGGAGG 89661			
QY	78	gggctgtgtgggaaggtctgggctggcagggacctgggctggggtctcttggccctgct 137			
Db	89662	GGGGGTGCAGGAAGAGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 89721			
QY	138	gtatgctttgcggcaatgctgggtgtgtgactctctggataacctggagatccctgcttt 197			
Db	89722	GTGTGCGGGTGGGGGGGGTGGGGAGGGAGGGAGGTGGGTGGGGGGGGGGTGTGGGGGG 89781			
QY	198	tggcgcaatccgggggttagttgtctatcaagactagagtggggtgagggaaagg 253			
Db	89782	AGGGGGGGGGGGGGGGGGGGTGGGAGGAGGGGGGGGGGGGGGGGGGGGGGG 89837			
RESULT	8				
AC095701/c		163525 bp	DNA	linear	HTG 20-DEC-2001
LOCUS		Rattus norvegicus clone CH230-9b18,	***	SEQUENCING IN PROGRESS	***,
DEFINITION		62 unordered pieces.			
AC095701					
VERSION		AC095701.2 GI:17942281			
KEYWORDS		HTG; HTGS, PHASE1.			
SOURCE		Norway rat.			
ORGANISM		Rattus norvegicus			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
		Rattus.			
		1 (bases 1 to 163525)			
		Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,			
		Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaris,J.,			
		Benton,J., Blinag,K., Blankenburg,K., Bonnin,D., Bouck,J.,			
		Bowen,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,			
		Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,			
		Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,			
		Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,			
		Coyle,M.D., Dathorne,R., David,R., Davila,M.L., Davis,C.,			
		Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,			
		Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,			
		Duggan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,			
		Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,			
		Foster,P., Francoz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,			
		Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,			
		Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,			
		Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,			
		Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,			
		Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,			
		Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,			
		Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,			
		Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,			
		Louisged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,			
		Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,			
		Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,			
		Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,			
		Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,			
		Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,			
		Ogih,M., Okuwonu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B.,			
		Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.B.,			
		Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,			
		Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,			
		Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,			
		Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,			
		Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,			
		Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,			
		Wall,R., Wang,S., Ward-Moore,S., Warren,K., Washington,C.,			

Fri May 24 10:49:50 2002

us-09-696-686-51.rge

Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 148921)
 Worley, K.C.
 Direct Submission
 Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15625779.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GDKJ
 Center clone name: CH230-9F14
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329first call to
 findphraplist
 Consensus quality: 117844 bases at least Q40
 Consensus quality: 130606 bases at least Q30
 Consensus quality: 138533 bases at least Q20
 Estimated insert size: 78668; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 87 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3865: contig of 3865 bp in length
 3866 3965: gap of unknown length
 3966 6975: contig of 3010 bp in length
 6976 7075: gap of unknown length
 7076 10377: contig of 3302 bp in length
 10378 10478: gap of unknown length
 10479 13677: contig of 3200 bp in length
 13678 13777: gap of unknown length
 13778 17257: contig of 3480 bp in length
 17258 17357: gap of unknown length
 17358 18597: contig of 1240 bp in length
 18598 18697: gap of unknown length
 18698 20220: contig of 1523 bp in length
 20221 20320: gap of unknown length
 20321 23582: contig of 3262 bp in length
 23583 23682: gap of unknown length
 23683 25938: contig of 2256 bp in length
 25939 26038: gap of unknown length
 26039 28627: contig of 2589 bp in length
 28628 28727: gap of unknown length
 28728 29791: contig of 1064 bp in length
 29792 29891: gap of unknown length
 29892 31945: contig of 2054 bp in length
 31946 32045: gap of unknown length
 32046 35054: contig of 3009 bp in length
 35055 35154: gap of unknown length
 35155 37831: contig of 2677 bp in length
 37832 37931: gap of unknown length
 40056: contig of 2125 bp in length
 40057 40156: gap of unknown length
 40157 41954: contig of 1798 bp in length
 41955 42054: gap of unknown length
 42055 43452: contig of 1398 bp in length

43453 43552: gap of unknown length
 43553 45687: contig of 2135 bp in length
 45688 45787: gap of unknown length
 45788 47477: contig of 1690 bp in length
 47478 47577: gap of unknown length
 47578 49684: contig of 2107 bp in length
 49685 49784: gap of unknown length
 49785 51965: contig of 2181 bp in length
 51966 52065: gap of unknown length
 52066 53630: contig of 1565 bp in length
 53631 54996: gap of unknown length
 54997 55096: contig of 1266 bp in length
 55097 56160: gap of unknown length
 56161 56260: contig of 1064 bp in length
 56261 57386: gap of unknown length
 57387 57486: contig of 1126 bp in length
 57487 58795: gap of unknown length
 58796 58895: contig of 1309 bp in length
 58896 60547: gap of unknown length
 60548 63005: contig of 2358 bp in length
 63006 63105: gap of unknown length
 63106 65100: contig of 1995 bp in length
 65101 65200: gap of unknown length
 65201 66722: contig of 1522 bp in length
 66723 66822: gap of unknown length
 66823 69735: contig of 2913 bp in length
 69736 69835: gap of unknown length
 69836 71274: contig of 1439 bp in length
 71275 71374: gap of unknown length
 71375 72856: contig of 1482 bp in length
 72857 72956: gap of unknown length
 72957 74011: contig of 1055 bp in length
 74012 74111: gap of unknown length
 74112 75778: contig of 1667 bp in length
 75779 75878: gap of unknown length
 75879 77264: contig of 1386 bp in length
 77265 77364: gap of unknown length
 77365 79442: contig of 2278 bp in length
 79443 79643: gap of unknown length
 79644 81472: contig of 1730 bp in length
 81473 81572: gap of unknown length
 81573 83102: contig of 1530 bp in length
 83103 83202: gap of unknown length
 83203 84298: contig of 1096 bp in length
 84299 84398: gap of unknown length
 84399 85406: contig of 1008 bp in length
 85407 85506: gap of unknown length
 85507 86520: contig of 1014 bp in length
 86521 86620: gap of unknown length
 86621 87692: contig of 1072 bp in length
 87693 87792: gap of unknown length
 87793 89176: contig of 1384 bp in length
 89177 89276: gap of unknown length
 89277 90575: contig of 1299 bp in length
 90576 91943: gap of unknown length
 91944 92043: contig of 1268 bp in length
 92044 93415: gap of unknown length
 93416 93515: gap of unknown length
 93516 94544: contig of 1029 bp in length
 94545 94644: gap of unknown length
 94645 95993: contig of 1349 bp in length
 95994 96093: gap of unknown length
 96094 97263: contig of 1070 bp in length
 97264 97263: gap of unknown length
 97264 99162: contig of 1899 bp in length
 99163 99262: gap of unknown length
 99263 100319: contig of 1057 bp in length
 100320 100419: gap of unknown length
 100420 102042: contig of 1623 bp in length
 102043 102142: gap of unknown length

Fri May 24 10:49:50 2002

```

* 28187 29106: contig of 920 bp in length
* 29107 29206: gap of 100 bp
* 29207 30153: contig of 949 bp in length
* 30156 30255: gap of 100 bp
* 30256 31182: contig of 927 bp in length
* 31183 31282: gap of 100 bp
* 31283 32108: contig of 826 bp in length
* 32109 32208: gap of 100 bp
* 32209 33133: contig of 925 bp in length
* 33134 33233: gap of 100 bp
* 33234 34221: contig of 988 bp in length
* 34222 34321: gap of 100 bp
* 34322 35205: contig of 884 bp in length
* 35206 35305: gap of 100 bp
* 35306 36272: contig of 967 bp in length
* 36273 36372: gap of 100 bp
* 36373 37360: contig of 988 bp in length
* 37361 37460: gap of 100 bp
* 37461 38362: contig of 902 bp in length
* 38363 38462: gap of 100 bp
* 38463 39425: contig of 963 bp in length
* 39426 39525: gap of 100 bp
* 39526 40391: contig of 866 bp in length
* 40392 40491: gap of 100 bp
* 40492 41461: contig of 970 bp in length
* 41462 42548: contig of 987 bp in length
* 42549 42648: gap of 100 bp
* 42649 43642: contig of 994 bp in length
* 43643 43742: gap of 100 bp
* 43743 44701: contig of 959 bp in length
* 44702 44801: gap of 100 bp
* 44802 45709: contig of 908 bp in length
* 45710 45809: gap of 100 bp
* 45810 46742: contig of 933 bp in length.

```

```

FEATURES
  source
    1. 46742
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="RP23-94114"
      /clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT 11493 a 8700 c 8777 g 12424 t 5348 others
ORIGIN

```

```

Query Match      12.98; Score 47.6; DB 2; Length 46742;
Best Local Similarity 30.4%; Pred. No. 0.44;
Matches 92; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY 13 gcaaatgcaaggtgagggcgccggcgccgtagccagcccgagggcgagg 72
Db 12752 GNNNGNNNNNGNGGGGNNNGGNNNGGNNNGGNNNGGNNNGGNNNGG 12811

QY 73 agggcgccgtgtggaaggtctggcctgagcagacctggctgggtctcttggcc 132
Db 12812 NGNGGGGGGGGNNNGNNNGNNNGGNNNGGNNNGGNNNGGNNNGGNNNG 12871

QY 133 ctgctgtgtcttgcgcaatgctggtgtgtgactctcgataacctggagatccct 192
Db 12872 NGNNNGGGGGNNNGNNNGGNNNGGNNNGNNNGNNNGNNNGNNNGNNNN 12931

QY 193 gcttttgagcgaatccggggtagttgtctatcaagactagagtgagggtgaggaag 252
Db 12932 GGNNNNNNGNNNGGNNNGGNNNGNNNNNNNNNNNGNNNGGNNNGGNN 12991

QY 253 gcttcaacagaagcctgctcgaatgaagattggccagggaaagcatgggtgcag 312
Db 12992 GNNNGGGGGGGGNNNGGNNNGGNNNGGNNNGGNNNGGNNNGGNNNGG 13051

QY 313 agg 315
Db 13052 GGG 13054

```

RESULT 11
AC093968/C
LOCUS
DEFINITION

AC093968
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

AC093968 199536 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus clone CH230-6J18, *** SEQUENCING IN PROGRESS ***,
57 unordered pieces.
AC093968
HTG: HTGS_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 199536)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T.R., Barbara,J.,
Benton,J., Bimege,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozando,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 199536)
Worley,K.C.

Submitted (13-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062167.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GBRK
Center clone name: CH230-6J18
----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findphraplist

Consensus quality: 175647 bases at least Q40
Consensus quality: 183306 bases at least Q30
Consensus quality: 188319 bases at least Q20
Estimated insert size: 181273; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.9x in Q20 bases; sum-of-contigs estimation

- * NOTE: Estimated insert size may differ from sequence length
(see http://www.hqsc.bcm.tmc.edu/docs/genbank_draft_data.html)
- * NOTE: This is a 'working draft' sequence. It currently
- * consists of 57 contigs. The true order of the pieces
- * is not known and their order in this sequence record is
- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will

```

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GCZP
Center clone name: CH230-8J16
----- Summary Statistics -----
Assembly program: Phrap: version 0.990329First call to
findPhrapList
Consensus quality: 73837 bases at least Q40
Consensus quality: 87577 bases at least Q30
Consensus quality: 95671 bases at least Q20
Estimated insert size: 65224; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarseqfp estimation
Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation
-----
* * NOTE: Estimated insert size may differ from sequence length
* * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* * NOTE: This is a 'working draft' sequence. It currently
* * consists of 60 contigs. The true order of the pieces
* * is not known and their order in this sequence record is
* * arbitrary. Gaps between the contigs are represented as
* * runs of N, but the exact sizes of the gaps are unknown.
* * This record will be updated with the finished sequence
* * as soon as it is available and the accession number will
* * be preserved.
* *
3530: contig of 3529 bp in length
3629: gap of unknown length
3630: contig of 3777 bp in length
7406: contig of 3777 bp in length
7506: gap of unknown length
10229: contig of 2723 bp in length
10329: gap of unknown length
12998: contig of 2669 bp in length
13030: contig of 2669 bp in length
13098: gap of unknown length
14893: contig of 1795 bp in length
14993: gap of unknown length
14894: contig of 2552 bp in length
14594: gap of unknown length
17546: gap of unknown length
17645: gap of unknown length
19280: contig of 1635 bp in length
19380: gap of unknown length
19281: contig of 2156 bp in length
19381: contig of 2156 bp in length
21537: gap of unknown length
21636: gap of unknown length
23581: contig of 1945 bp in length
23681: gap of unknown length
25799: contig of 2118 bp in length
25800: gap of unknown length
25899: gap of unknown length
28906: contig of 2907 bp in length
28907: gap of unknown length
28906: gap of unknown length
30583: contig of 1677 bp in length
30683: gap of unknown length
32640: contig of 1957 bp in length
32641: gap of unknown length
34130: contig of 1390 bp in length
34131: gap of unknown length
34231: gap of unknown length
35751: contig of 1521 bp in length
35851: gap of unknown length
37552: contig of 1754 bp in length
37606: gap of unknown length
37705: gap of unknown length
37706: gap of unknown length
40764: contig of 3059 bp in length
40765: gap of unknown length
40864: gap of unknown length
43662: contig of 2798 bp in length
43663: gap of unknown length
43663: gap of unknown length
44948: contig of 1186 bp in length
45048: gap of unknown length
45381: contig of 1333 bp in length
46382: gap of unknown length
46482: contig of 1043 bp in length
47524: gap of unknown length
47525: gap of unknown length
47624: gap of unknown length
47625: contig of 1375 bp in length
49000: gap of unknown length
49099: gap of unknown length
50960: contig of 1861 bp in length
50961: gap of unknown length
50961: gap of unknown length

```

[illegible]

[illegible]

9717 11129: contig of 1413 bp in length
 * 11130: gap of unknown length
 * 11230: contig of 1179 bp in length
 * 12409: gap of unknown length
 * 12509: contig of 1259 bp in length
 * 13767: contig of unknown length
 * 13867: gap of unknown length
 * 13868: 15044: contig of 1177 bp in length
 * 15045 15144: gap of unknown length
 * 15045 16325: contig of 1181 bp in length
 * 16326 16425: gap of unknown length
 * 16426 17578: contig of 1153 bp in length
 * 17579 17678: gap of unknown length
 * 17679 18845: contig of 1167 bp in length
 * 18846 18945: gap of unknown length
 * 18946 20060: contig of 1115 bp in length
 * 20061 20160: gap of unknown length
 * 20161 21258: contig of 1098 bp in length
 * 21259 21358: gap of unknown length
 * 21359 22390: contig of 1032 bp in length
 * 22391 22490: gap of unknown length
 * 22491 23599: contig of 1109 bp in length
 * 23600 23699: gap of unknown length
 * 23700 24907: contig of 1208 bp in length
 * 24908 25007: gap of unknown length
 * 25008 26121: contig of 1114 bp in length
 * 26122 26221: gap of unknown length
 * 26223 27273: contig of 1052 bp in length
 * 27274 27373: gap of unknown length
 * 27374 28472: contig of 1099 bp in length
 * 28473 28572: gap of unknown length
 * 28573 29683: contig of 1111 bp in length
 * 29684 29783: gap of unknown length
 * 29784 30834: contig of 1051 bp in length
 * 30835 30934: gap of unknown length
 * 30935 32135: contig of 1101 bp in length
 * 32136 33334: contig of 1199 bp in length
 * 33335 34334: gap of unknown length
 * 34335 34598: contig of 1164 bp in length
 * 34599 34698: gap of unknown length
 * 34699 35818: contig of 1120 bp in length
 * 35819 35918: gap of unknown length
 * 35920 37082: contig of 1164 bp in length
 * 37083 37182: gap of unknown length
 * 37183 38645: contig of 1463 bp in length
 * 38646 38745: gap of unknown length
 * 38746 39768: contig of 1023 bp in length
 * 39769 39868: gap of unknown length
 * 39869 40973: contig of 1105 bp in length
 * 40974 41073: gap of unknown length
 * 41074 42174: contig of 1101 bp in length
 * 42175 42274: gap of unknown length
 * 42275 43341: contig of 1067 bp in length
 * 43342 43441: gap of unknown length
 * 43443 44532: contig of 1091 bp in length
 * 44533 44632: gap of unknown length
 * 44633 46068: contig of 1436 bp in length
 * 46069 46168: gap of unknown length
 * 46169 47421: contig of 1253 bp in length
 * 47422 47521: gap of unknown length
 * 47522 48629: contig of 1108 bp in length
 * 48630 48729: gap of unknown length
 * 48730 49986: contig of 1257 bp in length
 * 49987 50086: gap of unknown length
 * 50087 51394: contig of 1308 bp in length
 * 51395 51494: gap of unknown length
 * 51495 53158: contig of 1664 bp in length
 * 53159 53258: gap of unknown length
 * 53259 54802: contig of 1544 bp in length
 * 54803 54902: gap of unknown length
 * 54903 56009: contig of 1107 bp in length
 * 56010 56109: gap of unknown length
 * 56110 57452: contig of 1343 bp in length

57453 57552: gap of unknown length
 * 57553 58667: contig of 1115 bp in length
 * 58668 58767: gap of unknown length
 * 58768 59924: contig of 1157 bp in length
 * 59925 60024: gap of unknown length
 * 60025 61159: contig of 1135 bp in length
 * 61160 62764: gap of unknown length
 * 62765 62864: gap of unknown length
 * 62865 64293: contig of 1429 bp in length
 * 64294 64393: gap of unknown length
 * 64394 65793: contig of 1400 bp in length
 * 65794 65893: gap of unknown length
 * 65894 67168: contig of 1275 bp in length
 * 67169 67268: gap of unknown length
 * 67269 69245: gap of unknown length
 * 69246 69345: gap of unknown length
 * 69346 70850: contig of 1505 bp in length
 * 70851 70950: gap of unknown length
 * 70951 73220: contig of 2270 bp in length
 * 73221 73320: gap of unknown length
 * 73321 76908: contig of 3588 bp in length
 * 76909 77008: gap of unknown length
 * 77009 78773: contig of 1765 bp in length
 * 78774 78873: gap of unknown length
 * 78874 81789: contig of 2916 bp in length
 * 81790 81889: gap of unknown length
 * 81890 84399: contig of 2510 bp in length
 * 84400 84499: gap of unknown length
 * 84500 87255: gap of unknown length
 * 87256 87355: gap of unknown length
 * 87356 91360: contig of 4005 bp in length
 * 91361 91460: gap of unknown length
 * 91461 95576: contig of 4116 bp in length
 * 95577 95676: gap of unknown length
 * 95677 99941: contig of 4265 bp in length
 * 99942 100041: gap of unknown length
 * 100042 105431: contig of 5390 bp in length
 * 105432 105531: gap of unknown length
 * 105532 112490: contig of 6959 bp in length
 * 112491 112590: gap of unknown length
 * 112591 121967: contig of 9377 bp in length
 * 121968 122067: gap of unknown length
 * 122068 132042: contig of 9975 bp in length
 * 132043 132142: gap of unknown length
 * 132143 143668: contig of 14226 bp in length
 * 143669 146369: gap of unknown length
 * 146370 156606: contig of 10138 bp in length
 * 156607 156706: gap of unknown length
 * 156707 166814: contig of 10108 bp in length
 * 166815 166914: gap of unknown length
 * 166915 177819: contig of 10905 bp in length
 * 177820 177919: gap of unknown length
 * 177920 191863: contig of 13944 bp in length
 * 191864 191963: gap of unknown length
 * 191964 204880: contig of 12917 bp in length
 * 204881 204980: gap of unknown length
 * 204981 220059: contig of 15079 bp in length
 * 220060 220159: gap of unknown length
 * 220160 238639: contig of 18480 bp in length
 * 238640 238739: gap of unknown length
 * 238740 262566: contig of 23827 bp in length
 * 262567 262666: gap of unknown length
 * 262667 300695: contig of 38029 bp in length.

FEATURES

source

1..300695="Mus musculus"

Query Match 12.8%; Score 47.2; DB 2; Length 300695;

Best Local Similarity 42.0%; Pred. No. 0.44;

Matches 97; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 19 gcgaagtgaggggcgccgcggcgtagcccaagcccgaggggcgggcg 78

Query Match 12.7%; Score 47; DB 2; Length 57787;
 Best Local Similarity 40.1%; Pred. No. 0.59; Indels 0; Gaps 0;
 Matches 101; Conservative

Qy 2 gcattgccagcaaatgcgaagtgcgaagtcgagggcgccggggcgccgagcctgcagcaagccgcga 61
 Db 26892 GGGGNGGNNNNNGGG 26833

Qy 62 gggcgaggagggcgccctgtggaagggtctgggctgcagcagcctggctggg 121
 Db 26832 GGGGNGGCGGG 26773

Qy 122 tctcttggccctctgtgtcttgcgcaatgcgtgggtgtgtgactctcgataacc 181
 Db 26772 GGGGNGGNNNNNGGG 26713

Qy 182 tggagatccctgcttttggcggaatccgggggtagtgtcatcaagactagagtgagg 241
 Db 26712 GGGGNGGNNNNNGGG 26653

Qy 242 gtgaggggaagg 253
 Db 26652 NNGGGGGGGGG 26641

RESULT 15
 AC094597/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-4A13, *** SEQUENCING IN PROGRESS ***
 AC094597
 AC094597
 AC094597.2 GI:17941361
 HTG: HTGS PHASE1
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 192182)
 Muzny D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,
 Alsbrooks S.L., Amaratunge H.C., Are J.R., Banks T., Barbara J.,
 Benton J., Binage K., Blankenburg K., Bonnin D., Bouck J.,
 Bowle S., Brieva M., Brown E., Brown M., Bryant N.P., Buhay C.,
 Burch P., Burkett C., Burrell K.B., Byrd N.C., Carron T.F.,
 Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,
 Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,
 Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
 Davy-Carroll L., Dederich D.A., Delaney K.R., Delgado O.,
 Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H.,
 Dugan-Rocha S., Durbin K.J., Earnhart C., Edgar D., Edwards C.C.,
 Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,
 Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,
 Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S.,
 Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A.,
 Hernandez J., Hernandez O., Hodgson A., Hoques M., Holloway C.,
 Hollins B., Honsi F., Howard S., Huber J., Hulyk S., Hume J.,
 Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S.,
 Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J.,
 Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,
 Lewis L., Li Z., Lichtarge O., Lieu C., Liu J., Liu W.,
 Louised H., Lozado R.J., Lu X., Lucier A., Lucier R., Luna R.,
 Ma J., Maheshwari M., Mapua P., Martin R., Martindale A.,
 Martinez E., Massey E., Mawhiney E., Mcleod M.P., Meador M.,
 Mei G., Metzker M., Miner G., Miner Z., Mitchell T., Mohabbat K.,
 Morgan M., Morris S., Moser M., Neal D., Newton S., Newton N.,
 Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokenkwo S.,
 Ouh M., Okwuonu G., Oragunye N., Oviado R., Pace A., Payton B.,
 Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L.,
 Quiles M., Ren Y., Rives M., Rojas A., Rojubokan I., Roife M.,
 Ruiz S., Savery G., Scherer S., Scott G., Shen H., Shoohtari N.,
 Sisson I., Sodergren E., Sonalke T., Sparks A., Stanley H.,
 Stone H., Sutton A., Svatek A., Tabor P., Tameris A., Tameris K.,
 Tang H., Tansey J., Taylor C., Taylor J., Telford B., Thomas N.,
 TONG H.

Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 192182)
 Worley, K.C.
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15624432.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GAZD
 Center clone name: CH230-4A13
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329 First call to
 findPhrapList
 Consensus quality: 135751 bases at least Q40
 Consensus quality: 146303 bases at least Q30
 Consensus quality: 153879 bases at least Q20
 Estimated insert size: 123383; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 94 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1
 * 5533: contig of 5533 bp in length
 * 5534: gap of unknown length
 * 5634: contig of 6404 bp in length
 * 12038: gap of unknown length
 * 12137: gap of unknown length
 * 17119: contig of 4982 bp in length
 * 17120: gap of unknown length
 * 22077: contig of 4858 bp in length
 * 22178: gap of unknown length
 * 25239: contig of 3062 bp in length
 * 25240: gap of unknown length
 * 30250: contig of 4911 bp in length
 * 30350: gap of unknown length
 * 33811: contig of 3461 bp in length
 * 33911: gap of unknown length
 * 37446: contig of 3535 bp in length
 * 37548: gap of unknown length
 * 37547: contig of 2816 bp in length
 * 40363: gap of unknown length
 * 40463: contig of 3075 bp in length
 * 43537: gap of unknown length
 * 43638: contig of 3241 bp in length
 * 46878: gap of unknown length
 * 46979: contig of 3376 bp in length
 * 50354: gap of unknown length
 * 50454: contig of 2235 bp in length
 * 52689: gap of unknown length
 * 52690: contig of 3642 bp in length
 * 56431: gap of unknown length
 * 56531: contig of 2675 bp in length
 * 59206: gap of unknown length
 * 59306: contig of 2200 bp in length
 * 59307: contig of 2200 bp in length
 *
 * 61507: gap of unknown length
 * 61607: contig of 3075 bp in length
 * 64682: gap of unknown length
 * 64782: contig of 2423 bp in length
 * 67205: gap of unknown length
 * 68555: contig of 1551 bp in length
 * 68856: gap of unknown length
 * 68956: contig of 3485 bp in length
 * 72441: gap of unknown length
 * 72541: contig of 1787 bp in length
 * 74328: gap of unknown length
 * 74428: contig of 3037 bp in length
 * 77465: gap of unknown length
 * 77565: contig of 1703 bp in length
 * 79268: gap of unknown length
 * 79368: contig of 2184 bp in length
 * 81552: gap of unknown length
 * 81652: contig of 1519 bp in length
 * 83171: gap of unknown length
 * 83271: contig of 1479 bp in length
 * 84750: gap of unknown length
 * 84850: contig of 1589 bp in length
 * 86439: gap of unknown length
 * 86539: contig of 1875 bp in length
 * 88414: gap of unknown length
 * 88514: contig of 1760 bp in length
 * 90274: gap of unknown length
 * 90374: contig of 2080 bp in length
 * 92454: gap of unknown length
 * 92554: contig of 2779 bp in length
 * 95333: gap of unknown length
 * 95433: contig of 1333 bp in length
 * 96765: gap of unknown length
 * 96865: contig of 2433 bp in length
 * 98866: gap of unknown length
 * 99299: contig of 1534 bp in length
 * 99399: gap of unknown length
 * 100933: contig of 1573 bp in length
 * 101033: gap of unknown length
 * 102606: contig of 1940 bp in length
 * 102706: gap of unknown length
 * 104646: contig of 1775 bp in length
 * 104746: gap of unknown length
 * 106521: contig of 1267 bp in length
 * 106621: gap of unknown length
 * 107888: contig of 1107 bp in length
 * 107988: gap of unknown length
 * 109095: contig of 2007 bp in length
 * 109195: gap of unknown length
 * 111202: contig of 2182 bp in length
 * 111302: gap of unknown length
 * 113484: contig of 1513 bp in length
 * 113584: gap of unknown length
 * 115097: contig of 2118 bp in length
 * 115197: gap of unknown length
 * 117315: contig of 1222 bp in length
 * 117415: gap of unknown length
 * 118637: contig of 2296 bp in length
 * 118737: gap of unknown length
 * 121033: contig of 1445 bp in length
 * 121133: gap of unknown length
 * 122578: contig of 2394 bp in length
 * 122678: gap of unknown length
 * 125072: contig of 1940 bp in length
 * 125172: gap of unknown length
 * 127112: contig of 1223 bp in length
 * 127212: gap of unknown length
 * 128435: contig of 1454 bp in length
 * 128535: gap of unknown length
 * 129988: contig of 1014 bp in length
 * 130089: gap of unknown length
 * 131103: contig of 1726 bp in length
 * 131203: gap of unknown length
 * 133028: contig of 1726 bp in length
 * 133929: gap of unknown length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

* 133029 134231: contig of 1203 bp in length
* 134232 134331: gap of unknown length
* 134332 135816: contig of 1485 bp in length
* 135817 135916: gap of unknown length
* 135917 137194: contig of 1278 bp in length
* 137195 137294: gap of unknown length
* 137295 139200: contig of 1906 bp in length
* 139201 139300: gap of unknown length

Query Match 12.7%; Score 47; DB 2; Length 192182;
Best Local Similarity 46.9%; Pred. NO. 0.51;
Matches 113; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 13 gcaaatgcgaagtgagggcgccgagggcgtagccaagcccgagggcgaggagg 72
Db 16776 GGAAGGA 16717

QY 73 gggcgggcctgtggaagggtctggcctggcagaccctgggctgggtctccttgccc 132
Db 16716 GGG 16657

QY 133 ctgctgtgtgttgaggcaatgctgggtgtgtgactctcgataaacctgagatccct 192
Db 16656 GGG 16597

QY 193 gcttttggcggaatccgggggtagtgtctcatcaagactagaggtggggtgaggggaag 252
Db 16596 GNNNGGG 16537

QY 253 g 253
Db 16536 G 16536

Search completed: May 24, 2002, 00:23:51
Job time: 11937 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 21:04:54 ; Search time 3953.4 Seconds
(without alignments)
132.333 Million cell updates/sec

Title: US-09-696-686-45

Perfect score: 25

Sequence: 1 agctcagacatgactccatgcc 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgtg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description

c	1	25	100.0	1882	10	BC013549	BC013549 Mus muscu
c	2	25	100.0	1909	10	AF361644	AF361644 Mus muscu
c	3	25	100.0	89855	2	AC074220	AC074220 Mus muscu
c	4	20.8	83.2	187142	2	AC097589	AC097589 Sus scrof
c	5	19.8	79.2	109963	2	AC107207	AC107207 Oryza sat
c	6	19.8	79.2	184288	2	AC011401	AC011401 Homo sapi
c	7	19.8	79.2	186780	9	AC005740	AC005740 Homo sapi
c	8	19.2	76.8	962	9	HSU80744	HSU80744 Homo sapien
c	9	19.2	76.8	1422	9	BC008898	BC008898 Homo sapi
c	10	19.2	76.8	1422	9	BC008961	BC008961 Homo sapi
c	11	19.2	76.8	1720	9	BC004423	BC004423 Homo sapi
c	12	19.2	76.8	18120	9	HSB20F6	HSB20F6 Human DNA s
c	13	19.2	76.8	113109	9	HS475N16	HS475N16 Human DNA
c	14	19.2	76.8	171247	9	AC007881	AC007881 Homo sapi
c	15	19.2	76.8	191719	2	AL671848	AL671848 Homo sapi
c	16	19.2	76.8	227238	2	AL669855	AL669855 Mus muscu
c	17	18.8	75.2	105686	9	AC012072	AC012072 Homo sapi
c	18	18.8	75.2	129947	9	AL355518	AL355518 Human DNA
c	19	18.8	75.2	170796	2	AC069416	AC069416 Homo sapi
c	20	18.6	75.2	182268	2	AC015591	AC015591 Homo sapi
c	21	18.6	74.4	643	9	HS1TAP1	HS1TAP1 Homo sapi
c	22	18.6	74.4	643	9	HS2TAP1	HS2TAP1 Homo sapi
c	23	18.6	74.4	1446	9	AF117647	AF117647 Homo sapi
c	24	18.6	74.4	1575	9	AB028645	AB028645 Homo sapi
c	25	18.6	74.4	1584	9	AF117646	AF117646 Homo sapi
c	26	18.6	74.4	3042	9	HS4430383	HS4430383 Homo sapi
c	27	18.6	74.4	3461	10	MMU430384	MMU430384 Mus muscu
c	28	18.6	74.4	9662	9	PTR251484	PTR251484 Pan trogl
c	29	18.6	74.4	9672	9	GG0251485	GG0251485 Gorilla g
c	30	18.6	74.4	11791	1	AE004763	AE004763 Pseudomon
c	31	18.6	74.4	27278	2	AC014009	AC014009 Drosophil
c	32	18.6	74.4	32941	1	SCM10	SCM10 Streptomy
c	33	18.6	74.4	37472	9	AC004539	AC004539 Homo sapi
c	34	18.6	74.4	37892	2	AC016540	AC016540 Homo sapi
c	35	18.6	74.4	66109	6	AX330551	AX330551 Sequence
c	36	18.6	74.4	66109	9	HSNHCAPG	HSNHCAPG H.sapiens g
c	37	18.6	74.4	80765	8	NCH1308	NCH1308 Neurospor
c	38	18.6	74.4	102677	2	AC095370	AC095370 Rattus no
c	39	18.6	74.4	106093	2	AC106270	AC106270 Rattus no
c	40	18.6	74.4	131906	2	AC083983	AC083983 Homo sapi
c	41	18.6	74.4	142215	2	AC092066	AC092066 Homo sapi
c	42	18.6	74.4	160757	2	AC087513	AC087513 Pan trogl
c	43	18.6	74.4	163183	3	AC009842	AC009842 Drosophil
c	44	18.6	74.4	171979	2	AC092516	AC092516 Papio cyn
c	45	18.6	74.4	173682	2	AC021316	AC021316 Homo sapi

ALIGNMENTS

RESULT 1

BC013549/c

LOCUS

DEFINITION

IMAGE:4222133, mRNA, complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC013549 Mus musculus, RIKEN cDNA 1600025D17 gene, clone MGC:19340 linear ROD 07-SEP-2001

IMAGE:4222133, mRNA, complete cds.

BC013549.1 GI:15488835

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 1882)

Strausberg, R.

Direct Submission

Submitted (04-SEP-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) md@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 25 Row: j Column: 9
 Location/Qualifiers

FEATURES
 source 1..1882
 /organism="Mus musculus"
 /db_xref="LocusID:72029"
 /db_xref="taxon:10090"
 /clone="MGC:19340 IMAGE:4222133"
 /tissue_type="Kidney, normal. 5 month old male mouse."
 /clone_lib="NCI CGAP_Kid14"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 91..921
 /codon_start=1
 /product="RIKEN cDNA 1600025D17 gene"
 /protein_id="AAH13549.1"
 /db_xref="GI:15488836"
 /translation="MFSMSLAPRCLLPFLPILLPLLLPAPKLGSPAGAEETDWR
 LPSKCEVCKYAVELKSAFETGKTEVDTGVLGDKGSGVKYKSDRLLEVTET
 ICKRLDYSLHRTGSGNRFKAGMSETFETLHNLVHKGVKVMYDIPYELWNETSAEVA
 DLKQCDVLVEFEFVEDWYRNHDEEDLFEFLCANHLVKGKDTSCLAERWSGKKGDI
 ASLGGKSKKRRSGVKGSSSGSKQRKELGLGEDANAEEGVQKASPLPHSPDPDL
 "

BASE COUNT 432 a 539 c 549 g 362 t
 ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1882;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatgactccatggccc 25
 |||||
 Db 110 AGCTCAGACATGGACTCCATGGCCC 86

RESULT 2
 AF361644/c 1909 bp mRNA linear ROD 12-MAY-2001
 LOCUS Mus musculus putative retinoic acid-regulated protein mRNA,
 complete cds.

DEFINITION
 ACCESSION AF361644
 VERSION AF361644.1 GI:14028766
 KEYWORDS house mouse.
 SOURCE Mus musculus

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 1909)
 Glozak, M.A., Li, Y., Reuille, R. and Rogers, M.B.
 Trapping and characterization of novel retinoic acid response
 elements

REFERENCE
 AUTHORS Glozak, M.A., Li, Y., Reuille, R. and Rogers, M.B.
 TITLE Direct Submission
 JOURNAL Submitted (15-MAR-2001) Biology, University of South Florida, 4202
 E. Fowler Avenue, Tampa, FL 33620, USA

Location/Qualifiers
 1..1909
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /cell_line="P19; derived from matings between C3H/He

FEATURES
 source 1..1909
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /cell_line="P19; derived from matings between C3H/He

females and male animals carrying an X chromosome derived
 from a feral mouse (McBurney & Rogers, Dev. Biol., 89,
 503)"
 /cell_type="embryonal carcinoma"
 /clone_lib="24 hr retinoic acid-treated P19 cDNA (Bouillet
 et al., Dev. Biol., 170, 420)"
 89..919
 /codon_start=1
 /product="putative retinoic acid-regulated protein"
 /protein_id="AAK52494.1"
 /db_xref="GI:14028767"
 /translation="MESMSLAPRCLLPFLPILLPLLLPAPKLGSPAGAEETDWR
 LPSKCEVCKYAVELKSAFETGKTEVDTGVLGDKGSGVKYKSDRLLEVTET
 ICKRLDYSLHRTGSGNRFKAGMSETFETLHNLVHKGVKVMYDIPYELWNETSAEVA
 DLKQCDVLVEFEFVEDWYRNHDEEDLFEFLCANHLVKGKDTSCLAERWSGKKGDI
 ASLGGKSKKRRSGVKGSSSGSKQRKELGLGEDANAEEGVQKASPLPHSPDPDL
 "

BASE COUNT 424 a 536 c 558 g 386 t 5 others
 ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1909;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatgactccatggccc 25
 |||||
 Db 108 AGCTCAGACATGGACTCCATGGCCC 84

RESULT 3
 AC074220 89855 bp DNA linear HTG 25-JAN-2002
 LOCUS Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
 WORKING DRAFT SEQUENCE, 38 unordered pieces.

DEFINITION
 ACCESSION AC074220.7 GI:13876390
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS house mouse.
 SOURCE Mus musculus

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 89855)
 Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
 Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
 Gordon, M., Goltz, J.S. and Kucherlapati, R.
 High Throughput Mouse Sequencing

REFERENCE
 AUTHORS Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
 Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
 Gordon, M., Goltz, J.S. and Kucherlapati, R.
 Direct Submission
 Submitted (20-JUL-2000) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

On Apr 28, 2001 this sequence version replaced gi:11276104.
 -----Genome Center
 Center: Harvard Partners Genome Center
 Center Code: HPGC
 Web site:
<http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mouse.html>
 Contact: gnkrm@capecod.bwh.harvard.edu
 e.html
 -----Summary Statistics
 Center project name: ADY
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100%
 *Consensus quality: 79355 at least Q20
 *Consensus quality: 75365 at least Q30
 *Consensus quality: 69022 at least Q40
 Estimated insert size: agarose-PP - N/A
 **Estimated insert size: 89115 - sum-of-contigs

Quality coverage: agarose-pp - N/A
Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs, the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 7962: contig of 7962 bp in length
* 7963 7982: gap of unknown length
* 7983 1324: contig of 5942 bp in length
* 1325 13944: gap of unknown length
* 13945 18990: contig of 5046 bp in length
* 18991 19010: gap of unknown length
* 19011 22501: contig of 3491 bp in length
* 22502 22521: gap of unknown length
* 22522 27077: contig of 4556 bp in length
* 27078 27097: gap of unknown length
* 27098 28819: contig of 1722 bp in length
* 28820 28839: gap of unknown length
* 28840 31224: contig of 2385 bp in length
* 31225 31244: gap of unknown length
* 31245 35093: contig of 3849 bp in length
* 35094 35113: gap of unknown length
* 35114 38611: contig of 3498 bp in length
* 38612 38632: contig of 3630 bp in length
* 38633 42261: contig of 3630 bp in length
* 42262 42282: gap of unknown length
* 42283 44130: contig of 1849 bp in length
* 44131 44150: gap of unknown length
* 44151 47059: contig of 2909 bp in length
* 47060 47079: gap of unknown length
* 47080 49496: contig of 2417 bp in length
* 49497 49516: gap of unknown length
* 49517 52237: contig of 2720 bp in length
* 52238 52256: gap of unknown length
* 52257 54270: contig of 2014 bp in length
* 54271 54290: gap of unknown length
* 54291 56271: contig of 1981 bp in length
* 56272 56291: gap of unknown length
* 56292 58849: contig of 2558 bp in length
* 58850 58869: gap of unknown length
* 58870 60621: contig of 1752 bp in length
* 60622 60641: gap of unknown length
* 60642 62255: contig of 1614 bp in length
* 62256 62276: gap of unknown length
* 62277 63085: contig of 810 bp in length
* 63086 63105: gap of unknown length
* 63106 65440: contig of 2335 bp in length
* 65441 65460: gap of unknown length
* 65461 67629: contig of 2169 bp in length
* 67630 67649: gap of unknown length
* 67650 69178: contig of 1529 bp in length
* 69179 69198: gap of unknown length
* 69199 70095: contig of 897 bp in length
* 70096 70115: gap of unknown length
* 70116 71821: contig of 1706 bp in length
* 71822 71841: gap of unknown length
* 71842 73506: contig of 1665 bp in length
* 73507 73526: gap of unknown length
* 73527 74749: contig of 1223 bp in length
* 74750 74769: gap of unknown length
* 74770 75963: contig of 1194 bp in length
* 75964 75983: gap of unknown length
* 75984 77905: contig of 1922 bp in length
* 77906 77925: gap of unknown length
* 77926 79236: contig of 1311 bp in length
* 79237 79256: gap of unknown length
* 79257 81332: contig of 2076 bp in length
* 81333 81352: gap of unknown length

* 81353 83169: contig of 1817 bp in length
* 83170 83189: gap of unknown length
* 83190 84096: contig of 907 bp in length
* 84097 84116: gap of unknown length
* 84117 85749: contig of 1633 bp in length
* 85750 85769: gap of unknown length
* 85770 86770: contig of 1001 bp in length
* 86771 86790: gap of unknown length
* 86791 88207: contig of 1417 bp in length
* 88208 88227: gap of unknown length
* 88228 88976: contig of 749 bp in length
* 88977 88996: gap of unknown length
* 88997 89855: contig of 859 bp in length.
* 88997 Location/Qualifiers
* 1. .89855 .
* /organism="Mus musculus"
* /strain="C57BL/6J"
* /db_xref="taxon:10090"
* /chromosome="17"
* /clone="RP23-76116"
* /sex="male"
* 1. .7962
* /note="assembly_name:Contig145"
* 7983. .13924
* /note="assembly_name:Contig144"
* 13945. .18990
* /note="assembly_name:Contig143"
* 19011. .22501
* /note="assembly_name:Contig142"
* 22522. .27077
* /note="assembly_name:Contig141"
* 27098. .28819
* /note="assembly_name:Contig140"
* 28840. .31224
* /note="assembly_name:Contig139"
* 31245. .35093
* /note="assembly_name:Contig138"
* 35114. .38611
* /note="assembly_name:Contig137"
* 38632. .42261
* /note="assembly_name:Contig136"
* 42282. .44130
* /note="assembly_name:Contig135"
* 44151. .47059
* /note="assembly_name:Contig134"
* 47080. .49496
* /note="assembly_name:Contig133"
* 49517. .52236
* /note="assembly_name:Contig132"
* 52257. .54270
* /note="assembly_name:Contig131"
* 54291. .56271
* /note="assembly_name:Contig130"
* 56292. .58849
* /note="assembly_name:Contig129"
* 58870. .60621
* /note="assembly_name:Contig128"
* 60642. .62255
* /note="assembly_name:Contig127"
* 62276. .63085
* /note="assembly_name:Contig126"
* 63106. .65440
* /note="assembly_name:Contig125"
* 65461. .67629
* /note="assembly_name:Contig124"
* 67650. .69178
* /note="assembly_name:Contig123"
* 69199. .70095
* /note="assembly_name:Contig122"
* 70116. .71821
* /note="assembly_name:Contig121"
* 71842. .73506
* /note="assembly_name:Contig120"

73527..74749 /note="assembly_name:Contig119"
74770..75963 /note="assembly_name:Contig118"
75984..77905 /note="assembly_name:Contig117"
77926..79236 /note="assembly_name:Contig116"
79257..81332 /note="assembly_name:Contig115"
81353..83169 /note="assembly_name:Contig114"
83190..84096 /note="assembly_name:Contig113"
100.04; Score 25; DB 2; Length 89855;
Best Local Similarity 100.04; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agctcagacatgactccatggccc 25
DB 6005 AGCTCAGACATGACTCCATGGCCC 6029
RESULT 4
AC097589 AC097589 187142 bp DNA linear HTG 19-OCT-2001
DEFINITION Sus scrofa clone RP44-78J22, WORKING DRAFT SEQUENCE, 4 unordered
pieces.
ACCESSION AC097589
VERSION AC097589.1 GI:16271927
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 187142)
Akhter.N., Ayele.K., Beckstrom-Sternberg,S.M., Benjamin.B.,
Blakesley.R.W., Bouffard,G.G., Breen.K., Brinkley.C., Brooks,S.,
Dietrich,N.L., Granite.S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-O.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.;
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 187142)
Green,E.D.
Direct Submission
Submitted (19-OCT-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
----- Project Information
Center project name: eth
Center clone name: 078J22
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 185160 bases at least Q40
Consensus quality: 185744 bases at least Q30
Consensus quality: 186130 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 186842; sum-of-contigs
Quality coverage: 10.12x in Q20 bases; agarose-fp
Quality coverage: 9.70x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 15610: contig of 15610 bp in length
* 15611 15710: gap of unknown length
* 15711 62856: contig of 47146 bp in length
* 62857 62956: gap of unknown length
* 62957 107606: contig of 44650 bp in length
* 107607 107707: gap of unknown length
* 107707 187142: contig of 79436 bp in length.
FEATURES
source 1..187142
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="RP44-78J22"
/clone_lib="RP44"
1..15610
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
15711..62856
/note="assembly_fragment"
62957..107606
/note="assembly_fragment"
clone_end:T7
vector_side:left"
107707..187142
/note="assembly_fragment"
53847 a 40777 c 40348 g 51863 t 307 others
BASE COUNT
ORIGIN
Query Match 83.2%; Score 20.8; DB 2; Length 187142;
Best Local Similarity 91.7%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 gctcagacatgactccatggccc 25
DB 112578 GATCAGAATGGACTCCATGGCCC 112601
RESULT 5
AC107207/c AC107207 109963 bp DNA linear HTG 06-FEB-2002
LOCUS
DEFINITION Oryza sativa chromosome 3 clone OSJNBb0106M04, *** SEQUENCING IN
PROGRESS ***, 7 unordered pieces.
ACCESSION AC107207
VERSION AC107207.2 GI:18543002
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoidae; Oryzaceae; Oryza.
1 (bases 1 to 109963)
Buell,R., Hsiao,J., Zismann,V., Moffat,K.M., Hill,J.,
Gansberger,K., Burgess,S., Jarrahi,B., Shvartsbeyn,M., Brenner,M.,
Ciecko,A., Pai,G., Vanaken,S., Hansen,C., Utterbach,T.,
Feldblyum,T., Khalak,H.G., Yuan,Q., Quackenbush,J., White,O.,
Salzberg,S. and Fraser,C.
Oryza sativa ssp. japonica cv. Nipponbare OSJNBb0106M04 BAC genomic
sequence
Unpublished
2 (bases 1 to 109963)
Buell,R.
Direct Submission
Submitted (16-JAN-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Feb 6, 2002 this sequence version replaced gi:18158373.
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 21690: contig of 21690 bp in length
 * 21691 21732: gap of unknown length
 * 21733 63992: contig of 42260 bp in length
 * 63993 64034: gap of unknown length
 * 64035 73846: contig of 9812 bp in length
 * 73847 73888: gap of unknown length
 * 73889 89053: contig of 15165 bp in length
 * 89054 89095: gap of unknown length
 * 89096 97762: contig of 8667 bp in length
 * 97763 97803: gap of unknown length
 * 97804 100038: contig of 2235 bp in length
 * 100039 100079: gap of unknown length
 * 100080 109963: contig of 9884 bp in length.

FEATURES

Location/Qualifiers

1..109963
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="3"
 /clone="OSJNB0106M04"

BASE COUNT 30639 a 24707 c 24423 g 29906 t 288 others

ORIGIN

Query Match 79.2% Score 19.8; DB 2; Length 109963;
 Best Local Similarity 91.3%; Pred. No. 60;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gctcagacatggactccatgcc 24

Db 34022 GTTCAGATGGACTGCATGCC 34000

RESULT

AC011401 6
 LOCUS Homo sapiens chromosome 5 clone CTB-35K5, linear HTG 04-OCT-2001
 DEFINITION 11 ordered pieces.

AC011401
 AC011401.8 GI:12830130
 HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 184268)
 DOE Joint Genome Institute.

JOURNAL Sequencing of Human Chromosome 5
 Unpublished

REFERENCE 2 (bases 1 to 184268)
 DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Feb 14, 2001 this sequence version replaced gi:9954634.

COMMENT

-----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

 Project Information
 Center Project Name: 77889, H187
 Center clone name: C1978SKB_35K5

Summary Statistics

Consensus quality: 173129 bases at least Q40
 Consensus quality: 180377 bases at least Q30
 Consensus quality: 182040 bases at least Q20
 Estimated insert size: 200000; pulse field gel estimation
 Estimated insert size: 183268; sum-of-contigs estimation
 Quality coverage: 6.62 in Q20 bases; pulse field gel estimation
 Quality coverage: 7.22 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 21632: contig of 21632 bp in length
 * 21732 21733: gap of unknown length
 * 21733 29338: contig of 7606 bp in length
 * 29339 29438: gap of unknown length
 * 29439 32479: contig of 2941 bp in length
 * 32480 32479: gap of unknown length
 * 32480 47614: contig of 15135 bp in length
 * 47615 47715: gap of unknown length
 * 47715 63279: contig of 15565 bp in length
 * 63280 63379: gap of unknown length
 * 63380 69838: contig of 6459 bp in length
 * 69839 69938: gap of unknown length
 * 69939 71512: contig of 1574 bp in length
 * 71513 71612: gap of unknown length
 * 71613 77519: contig of 5907 bp in length
 * 77520 77619: gap of unknown length
 * 77620 119849: contig of 42230 bp in length
 * 119850 119949: gap of unknown length
 * 119950 146636: contig of 26687 bp in length
 * 146637 146736: gap of unknown length
 * 146737 184268: contig of 37532 bp in length.

FEATURES

source

1..184268
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTB-35K5"
 /clone_lib="CalTech human BAC library B"
 BASE COUNT 52948 a 40258 c 38925 g 51136 t 1001 others

ORIGIN

Query Match 79.2% Score 19.8; DB 2; Length 184268;
 Best Local Similarity 91.3%; Pred. No. 61;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gctcagacatggactccatgcc 24

Db 9417 GTTCAGAAATGGACTCCATGCC 9439

RESULT

AC005740 7
 LOCUS AC005740/c

DEFINITION

Homo sapiens chromosome 5p, BAC clone 50g21 (LBNL H154), complete
 sequence.

ACCESSION

AC005740

VERSION

AC005740.1 GI:3687210

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 186780)
 Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
 Kadner, K., Miguel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

Sequencing of human chromosome 5

JOURNAL
REFERENCE
2 (bases 1 to 186780)

Unpublished
Ricke, D.O.

JOURNAL
REFERENCE
3 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
4 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
5 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
6 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
7 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
8 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
9 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
10 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
11 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
12 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
13 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
14 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
15 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
16 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
17 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
18 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
19 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
20 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
21 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

JOURNAL
REFERENCE
22 (bases 1 to 186780)

Unpublished
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Adams, K., Miquel, T., Miller, C., Pittluck, S., Pollard, M.,

Rojeski, H., Subramanian, S. and Martin, C.H.

/rpt_unit-A
complement(12372..12811)
/rpt_family="L1"
12900..13016
/rpt_family="Alu"
complement(13084..13169)
/rpt_family="L1"
complement(13597..13889)
/rpt_family="Alu"
complement(15617..15931)
/rpt_family="Alu"
15793..15817
/note="(T)25"
/rpt_type=tandem
/rpt_unit=T
complement(16185..16462)
/rpt_family="Alu"
complement(16604..16902)
/rpt_family="Alu"
17036..17206
/rpt_family="MLT1"
17220..17832
/rpt_family="Alu"
17626..17649
/note="(A)24"
/rpt_type=tandem
/rpt_unit=A
17874..18004
/rpt_family="MLT1"
18255..18529
/rpt_family="Alu"
complement(18691..18870)
/rpt_family="MER3"
complement(18788..18822)
/rpt_family="MER33"
complement(18983..19265)
/rpt_family="Alu"
20976..21209
/rpt_family="Alu"
21298..21414
/rpt_family="L1"
complement(21413..21577)
/rpt_family="Alu"
complement(21730..22011)
/rpt_family="Alu"
22064..22271
/rpt_family="Alu"
22531..22595
/note="GRAIL 2 excellent exon, frame 2"
22780..23054
/rpt_family="Alu"
23190..23489
/rpt_family="Alu"
23849..24002
/rpt_family="MER39B"
complement(24000..26149)
/rpt_family="L1"
26149..27398
/rpt_family="L1"
27432..27706
/rpt_family="Alu"
27721..28129
/rpt_family="MER39"
28498..28774
/rpt_family="Alu"
complement(29403..29490)
/rpt_family="MIR2"
29525..29716
/rpt_family="Alu"
30196..30275
/rpt_family="MIR"
complement(30377..30662)
/rpt_family="Alu"

<hr/>					
FEATURES	Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA				
source	Location/Qualifiers				
	1..962				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/tissue_type="cerebral cortex"				
gene	1..962				
	/gene="CTG4a"				
CDS	388..819				
	/gene="CTG4a"				
	/note="polyucleine rich"				
	/codon_start=1				
	/product="CTG4a"				
	/protein_id="AA891442.1"				
	/db_xref="GI:2565063"				
	/translation="MDSMPFASRCLLLLPLLLLLLLPAPELGPSQAGSENDWVR LPSCVECKYVAVELKSAFETGKTKEVIGTGYILDQKASGVKYTKSDLRLIETEVTR ICKRLLIIACTRRGPAAIDLPACQRPLRHYYTWYTKGSRW"				
repeat_region	439..462				
	/rpt_type=tandem				
	/rpt_unit=CTG				
BASE COUNT	200 a 252 c 315 g 194 t 1 others				
ORIGIN					
<hr/>					
Query Match	76.8%; Score 19.2; DB 9; Length 962;				
Best Local Similarity	87.5%; Pred. No. le+02;				
Matches	21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;				
<hr/>					
Qy	2 gctcagacatggactccatgcccc 25				
Dd	406 GCTCAGGCATTGAATCCATGCCCC 383				
<hr/>					
RESULT	9				
BC008898/c					
LOCUS	BC008898 1422 bp mRNA linear PRI 12-JUL-2001				
DEFINITION	Homo sapiens, clone MGC:1220 IMAGE:2959532, mRNA, complete cds.				
ACCESSION	BC008898				
VERSION	BC008898.1 GI:14286203				
KEYWORDS	MGC.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1422) Direct Submission Strausberg,R. Submitted (29-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1IA03, Bethesda, MD 20892-2590, USA				
REMARK	NTF-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk Email: cgabs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Institute for Systems Biology http://www.systemsbio.org contact: amadan@systemsbiology.org Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettman and Anuradha Madan				
<hr/>					
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://Image.llnl.gov Series: IRAL Plate: 2 Row: m Column: 12 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6002306.					
FEATURES	Location/Qualifiers				
source	1..1422				
	/organism="Homo sapiens"				

repeat_region	complement(30672..31137)	/rpt_family="L1"
repeat_region	31138..31564	/rpt_family="MERS7B"
repeat_region	complement(31565..31637)	/rpt_family="L1"
misc_feature	join(31640..31733,31791..32094)	/note="85% & 91% identity EST zb50f04.r1"
misc_feature	db_xref="dbEST:W24400"	31917..32270
repeat_region	/note="90% identity EST z104c10.r1"	32354..32655
repeat_region	/rpt_family="Alu"	complement(32690..32798)
repeat_region	/rpt_family="Alu"	complement(32793..32997)
repeat_region	/rpt_family="L1"	33000..33296
repeat_region	/rpt_family="Alu"	34147..34354
repeat_region	/rpt_family="Alu"	complement(34383..34822)
repeat_region	/rpt_family="L1"	35142..35273
repeat_region	/rpt_family="MER87"	36021..36304
repeat_region	/rpt_family="Alu"	36482..36594
misc_feature	/rpt_family="MIR"	36830..36949
repeat_region	/note="GRAIL 2 excellent exon, frame 1"	complement(37800..37874)
repeat_region	/rpt_family="MER91B"	38633..38910
repeat_region	/rpt_family="Alu"	38909..38935
repeat_region	/note="(AAT)9"	/rpt_type=tandem
repeat_region	/rpt_unit=AAT	complement(39679..39973)
Query Match	79.2%; Score 19.8; DB 9; Length 186780;	
Best Local Similarity	91.3%; Pred. No. 61;	
Matches 21; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
Qy 2	gctcagacatggactccatgcc 24	
Db 47259	GTTCAGAAATGGACTCCATGCC 47237	
RESULT 8		
HSU80744/c	962 bp mRNA linear	PRI 18-DEC-1997
LOCUS		
DEFINITION	Homo sapiens CTG4a mRNA, complete cds.	
ACCESSION	U80744	
VERSION	U80744.1	GI:2565062
KEYWORDS		
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 962)	
TITLE	Marcolis, R. L., Abraham, M. R., Gatchell, S. B., Li, S. H., Kidwai, A. S., Breschel, T. S., Stine, O. C., Callahan, C., McInnis, M. G. and Ross, C. A.	
JOURNAL	CONAS with long CAG trinucleotide repeats from human brain	
MEDLINE	Hum. Genet. 100 (1), 114-122 (1997)	
REFERENCE	97369492	
AUTHORS	2 (bases 1 to 962)	
TITLE	Marcolis, R. L., Abraham, M. R., Gatchell, S. B., Li, S. H., Kidwai, A. S., Breschel, T. S., Stine, O. C., Callahan, C., McInnis, M. G. and Ross, C. A.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of	

```

/db_xref="taxon:9606"
/clone="MGC:1220 IMAGE:2959532"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_15"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
78. .914
/codon_start=1
/product="Unknown (protein for MGC:1220)"
/protein_id="AAH08896.1"
/db_xref="GI:14286204"
/translation="MDSMPFASRCRLLLPLLLLLLLPAPELGPSQAGAEENDWVR
LPSCVECKYVAVELKSAFETGKTKEVIGTGILDOKASGVKTKSDRLIEVTET
ICKRLDYSILHKERTGNSRPAKGMSETFTLHNLVHKGVKVMIDPYELWNETSAEVA
DLKKQCDVLVEEFEEVIEDWVRNHOEDLTFLCANHLVKGKDTSCLAEQWSGKKGDT
AALGGKKKKSSRAKAGRSSSSKORKELGLEGDPSPDEDEGIQKASPLTHSPDP
EL"
BASE COUNT      351 a  400 c  422 g  249 t
ORIGIN

Query Match      76.8%; Score 19.2; DB 9; Length 1422;
Best Local Similarity 87.5%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gctcagacatgactccatggccc 25
||||| ||| || ||||| |||||
Db 96 GCTCAGGCATTGAATCCATGGCCC 73

RESULT 10
BC008961/c
LOCUS      1422 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone MGC:4122 IMAGE:2959532, mRNA, complete cds.
ACCESSION BC008961
VERSION BC008961.1 GI:14286319
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1422)
Direct Submission
Strausberg,R.
Submitted (29-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: anadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 10 Row: m Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6002306.
Location/Qualifiers
1. .1422
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:4122 IMAGE:2959532"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_15"
/lab_host="DH10B-R"

FEATURES
source

```

```

/note="Vector: pOTB7"
78. .914
/codon_start=1
/product="Unknown (protein for MGC:4122)"
/protein_id="AAH08961.1"
/db_xref="GI:14286320"
/translation="MDSMPFASRCRLLLPLLLLLLLPAPELGPSQAGAEENDWVR
LPSCVECKYVAVELKSAFETGKTKEVIGTGILDOKASGVKTKSDRLIEVTET
ICKRLDYSILHKERTGNSRPAKGMSETFTLHNLVHKGVKVMIDPYELWNETSAEVA
DLKKQCDVLVEEFEEVIEDWVRNHOEDLTFLCANHLVKGKDTSCLAEQWSGKKGDT
AALGGKKKKSSRAKAGRSSSSKORKELGLEGDPSPDEDEGIQKASPLTHSPDP
EL"
BASE COUNT      351 a  400 c  422 g  249 t
ORIGIN

Query Match      76.8%; Score 19.2; DB 9; Length 1422;
Best Local Similarity 87.5%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gctcagacatgactccatggccc 25
||||| ||| || ||||| |||||
Db 96 GCTCAGGCATTGAATCCATGGCCC 73

RESULT 11
BC004423/c
LOCUS      1720 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone MGC:3530 IMAGE:2819660, mRNA, complete cds.
ACCESSION BC004423
VERSION BC004423.1 GI:13325207
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1720)
Direct Submission
Strausberg,R.
Submitted (12-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/BTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranasa Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 9 Row: o Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5729791.
Location/Qualifiers
1. 1720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:3530 IMAGE:2819660"
/tissue_type="lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"

FEATURES
source

```

```
/lab_host="DH10B-R"
/note="Vector: pOTB7"
388. 1224
/codon_start=1
/product="Unknown (protein for MGC:3530)"
/protein_id="AAH04423.1"
/db_xref="GI:13325208"
/translation="MDSMPAPASRCLLLPLLLLLLLLPAPELGPQAGAEENDWVR
LPKSCVCKYVAVELKSAPECTKTEVIGTGYGILDDOKAGVKYTKSDLRLIEVTET
ICKRLDYSLHKERTGSRNRAKMSSEFFELHNLVHKGYKVVMDIPELWNETSAEVA
DLKKQCDVLEVEEVEDIYRNHQBEDLFEFCANHLVKGKDTSCLAEGWSKKGGDT
AALGGKKKKKSRRAAGRSSSKQKELGLEGDPSPDEGIGOKASPLTHSPDD
EL"
BASE COUNT 389 a 491 c 526 g 314 t
ORIGIN
Query Match 76.8%; Score 19.2; DB 9; Length 1720;
Best Local Similarity 87.5%; Pred. No. 1.1e+07;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 gctcagatggactccatggccc 25
||||| ||| || ||||| |||||
Db 406 GCTCAGGATTGAATCCATGGCCC 383

RESULT 12
LOCUS HSB20F6 18120 bp DNA linear PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone SC22CB-20F6 on chromosome 22,
complete sequence.
ACCESSION Z82174
VERSION Z82174.2 GI:6572311
KEYWORDS HFG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Whiteley, M. Isson
Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:1695140.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
SC22CB-20F6 is from the human chromosome 22-specific cosmid library
(SC22CB) constructed at the Sanger Centre by Mark Ross and Cordelia
Langford.
VECTOR: lawlist16
IMPORTANT: This sequence is not the entire insert of clone
SC22CB-20F6 it may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
```

```
between neighbouring submissions.
The true right end of clone SC22CB-20F6 is at 18120 in this
sequence. The true right end of clone SC22CB-79B4 is at 105 in this
sequence.
FEATURES
source
1. 18120
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="SC22CB-20F6"
/clone_lib="SC22cB"
3. 189
/note="AluJb repeat: matches 109. .281 of consensus"
665. 792
/note="FLAM_C repeat: matches 1. .131 of consensus"
820. 1366
/note="L1PB1 repeat: matches 5634. .6155 of consensus"
858. 862
/note="uncertain number of bases"
1437. 1472
/note="L2 repeat: matches 2704. .2739 of consensus"
1758. 1950
/note="L2 repeat: matches 2539. .2740 of consensus"
1942. 2100
/note="MIR repeat: matches 48. .201 of consensus"
2186. 2208
/note="L2 repeat: matches 2597. .2617 of consensus"
2209. 2478
/note="AluJo repeat: matches 1. .302 of consensus"
2479. 2599
/note="L2 repeat: matches 2617. .2749 of consensus"
3589. 3718
/note="MIR repeat: matches 47. .184 of consensus"
4853. 5165
/note="Alusq repeat: matches 1. .313 of consensus"
5313. 5417
/note="MIR repeat: matches 45. .155 of consensus"
5507. 5629
/note="MIR repeat: matches 16. .140 of consensus"
6183. 6226
/note="L1P repeat: matches 87. .130 of consensus"
7093. 7209
/note="MIR repeat: matches 50. .180 of consensus"
7772. 8079
/note="Alusx repeat: matches 1. .306 of consensus"
8203. 8445
/note="L2 repeat: matches 2396. .2750 of consensus"
8545. 8841
/note="Alusx repeat: matches 1. .293 of consensus"
10285. 10473
/note="MER20 repeat: matches 1. .218 of consensus"
10551. 10606
/note="28 copies 2 mer ac 75 conserved"
10651. 11166
/note="LMC1 repeat: matches 5770. .6315 of consensus"
12417. 12852
/note="MLT1C repeat: matches 1. .466 of consensus"
13073. 13314
/note="Alusq repeat: matches 21. .296 of consensus"
13496. 13500
/note="uncertain number of bases"
13590. 13667
/note="MIR repeat: matches 61. .140 of consensus"
13677. 13721
/note="MLT1C repeat: matches 1. .45 of consensus"
13834. 14027
/note="MLT1D repeat: matches 1. .204 of consensus"
14069. 14363
/note="Alusx repeat: matches 1. .293 of consensus"
14443. 15245
/note="match: GSS: Em:AQ747573"
15017. 15272
/note="Alusx repeat: matches 35. .290 of consensus"
```



```
EL"
15046. .15456
/Note="match: STS: Em:G37319"
complement(17091. .17133)
/gene="dJ475N16.1"
/Note="match: STS: Em:G25145 Em:G27441"
21516. .22025
/Note="CpG island"
/evidence=not_experimental
21749. .21784
/Note="match: STS: Em:AF021121"
complement(21753. .21788)
/gene="dJ475N16.1"
/Note="match: STS: Em:G09564"
21759. .21794
/Note="match: STS: Em:267474"
21762. .21797
/Note="match: STS: Em:G09799"
complement(21762. .21793)
/gene="dJ475N16.1"
/Note="match: STS: Em:G09544"
complement(22138. .22473)
/Note="match: GSS: Em:AQ756680"
23412. .23718
/Note="AluSc repeat: matches 2. .309 of consensus"
complement(23719. .23975)
/Note="match: GSS: Em:AQ267139"
complement(24876. .25553)
/Note="match: GSS: Em:AQ383009"
complement(25571)
/gene="dJ475N16.2"
complement(join(25571. .26188,27131. .27175,28057. .28377,
35277. .35438))
/gene="dJ475N16.2"
/Note="match: CDNAS: Em:U16958 Em:U36759 Em:U38996; match:
ESTs: Em:AA913164 Em:AI263200 Em:AI285999 Em:AW014894
Em:AI767176"
/product="dJ475N16.2 (pre-T cell receptor alpha)"
/evidence=not_experimental
complement(25571. .35438)
/gene="dJ475N16.2"
complement(25590. .25595)
/gene="dJ475N16.2"
complement(join(25722. .26188,27131. .27175,28057. .28377,
35277. .35334))
/gene="dJ475N16.2"
/codon_start=1
/evidence=not_experimental
/product="dJ475N16.2 (pre-T cell receptor alpha)"
/protein_id="CAB75302.1"
/db_xref="GI:6969164"
/translation="MAGTLLLLALGCPALPTGVGTPPPLAPPTMLLVGKQQWV
VCLVLDVAPGLDSPWFNSAGNSALDAFYGSPATDGTWNLHLPLSPSELASH
EPLVCHTGPAGHSRSTQPHLSEASTATCPQEPRLRGCGGLRAPERFLLAGTPG
GALWGLVRLLLFLDLLLCLSCDPCGAPLPSPATTTLRALGSHRLHPATETG
GREATSPQPPQDRRRGDTTPGKRKPGSPVMGEGLSSYPTPCPAQAWCSRSALRAP
SSLGAFAGDLPPLQAGAA"
29689. .29994
/Note="AluSc repeat: matches 1. .306 of consensus"
30593. .30668
/Note="L1MB8 repeat: matches 6101. .6170 of consensus"
33433. .33494
/Note="31 copies 2 mer tt 69% conserved"
38170. .38299
/Note="L1MB3 repeat: matches 6024. .6137 of consensus"
38610. .38858
/Note="L1MB3 repeat: matches 5711. .6024 of consensus"
38941. .39002
/Note="AluJ/FLAM repeat: matches 19. .80 of consensus"
39289. .39862
/Note="CpG island"
/evidence=not_experimental
40745. .40851

repeat_region
41336. .41377
/Note="L1P4 repeat: matches 5703. .5817 of consensus"
/Note="21 copies 2 mer ac 90% conserved"
41779. .41844

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 9; Length 113109;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gctcagacatggactccatgccc 25
||||| ||| || ||||| |||
Db 21815 GCTCAGGATTGAATCCATGCC 21838

RESULT 14
AC007881
LOCUS Homo sapiens chromosome 2 clone RP11-467P9, complete sequence.
DEFINITION AC007881
ACCESSION AC007881
VERSION AC007881.4 GI:11181854
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (19-JUN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (16-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Nov 16, 2000 this sequence version replaced gi:7631010.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: H_NH0467P09
-----
Location/Qualifiers
1. 171247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-467P9"
BASE COUNT 46831 a 37692 c 38499 g 48225 t
ORIGIN

Query Match
76.8%; Score 19.2; DB 9; Length 171247;
```

Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatgcc 24

Db 26029 AGCTGATACAGGACATCCATGCC 26052

RESULT 15

AL671848 191719 bp DNA linear HTG 08-FEB-2002
LOCUS Mus musculus chromosome X clone RP23-110K5, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.

AL671848

ACCESSION

AL671848.3 GI:18643855

KEYWORDS HTG; HTGS-PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (sites)

AUTHORS Burton.J.

TITLE Direct Submission

JOURNAL Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 10, 2002 this sequence version replaced gi:18491505.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: bm110K5

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 189652 bases at least Q40

Consensus quality: 190201 bases at least Q30

Consensus quality: 190528 bases at least Q20

Insert size: 191119; sum-of-contigs

Insert size: 179804; 0.5% error; agarose-fp

Quality coverage: 9.71x in Q20 bases; sum-of-contigs Quality

coverage: 10.44x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES

source

1. 191719

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="X"

/clone="RP23-110K5"

/clone_lib="RPCI-23"

1. 2127

/note="assembly_fragment:03655"

fragment_chain:1

clone_end:T7

vector_side:left"

2228. 40239

/note="assembly_fragment:04272"

fragment_chain:1"

40340. 64658

/note="assembly_fragment:02305"

fragment_chain:1"

64759. 81283

/note="assembly_fragment:04002"

fragment_chain:1"

81384. 140088

/note="assembly_fragment:03265"

fragment_chain:1"

misc_feature 140189. 157395
/note="assembly_fragment:04052"

fragment_chain:1"

157496. 191719

misc_feature /note="assembly_fragment:04056"

fragment_chain:1

clone_end:SP6

vector_side:right"

BASE COUNT 54268 a 37719 c 37756 g 61366 t 610 others

ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 191719;

Best Local Similarity 87.5%; Pred. No. 1.2e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatgcc 24

Db 68165 AGCTCAGAGATGTAGTCCATGCC 68188

Search completed: May 24, 2002, 00:12:11

Job time: 11237 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 23:05:27 ; Search time 4203.75 seconds
(without alignments)
80.267 Million cell updates/sec

Title: US-09-696-686-46
Perfect score: 25
Sequence: 1 tgcgattgccagcaaatgcgaagt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :					EST:*				
1: em_estba:*					2: em_esthum:*				
3: em_estin:*					4: em_estmu:*				
5: em_estov:*					6: em_estpl:*				
7: em_estro:*					8: em_htc:*				
9: gb_est1:*					10: gb_est2:*				
11: gb_htc:*					12: gb_gss:*				
13: em_gss_hum:*					14: em_gss_inv:*				
15: em_gss_pln:*					16: em_gss_vrt:*				

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	260	9	BB666889
2	25	100.0	330	9	BB856691
3	25	100.0	382	9	AA048794
4	25	100.0	406	9	AA002532
5	25	100.0	435	9	AI098985
6	25	100.0	472	9	BB863087
7	25	100.0	483	9	BB857917
8	25	100.0	485	9	BB857913
9	25	100.0	495	10	BG062828
10	25	100.0	502	9	BB853356
11	25	100.0	608	9	BB605812
12	25	100.0	623	9	BB661026
13	25	100.0	625	9	BB633565
14	25	100.0	641	9	BB637089
15	25	100.0	643	9	BB661603
16	25	100.0	649	9	BB652011
17	25	100.0	651	10	BE289316

18	25	100.0	725	10	BG296771	BG296771	602394543
19	25	100.0	741	10	BI081472	BI081472	602879994
20	25	100.0	776	10	BG864846	BG864846	602799277
21	25	100.0	779	10	BI688404	BI688404	603316338
22	25	100.0	790	10	BF301937	BF301937	602032966
23	25	100.0	793	10	BI104487	BI104487	602890091
24	25	100.0	1553	11	AK005532	AK005532	Mus muscu
25	24	96.0	490	9	BE226435	BE226435	la20q05.y
26	24	96.0	637	9	BB650071	BB650071	BB650071
27	24	96.0	821	10	BI653380	BI653380	Mus muscu
28	24	96.0	1586	11	AK010702	AK010702	Mus muscu
29	23.4	93.6	249	9	AW356648	AW356648	39152 MAR
30	23.4	93.6	305	10	BE484870	BE484870	171676 BA
31	23.4	93.6	341	9	AW446522	AW446522	86241 MAR
32	23.4	93.6	379	10	BM105433	BM105433	508902 MA
33	23.4	93.6	410	10	BM287749	BM287749	528547 MA
34	23.4	93.6	411	9	AW357963	AW357963	41782 MAR
35	23.4	93.6	416	10	BF651657	BF651657	274494 MA
36	23.4	93.6	429	9	AW669666	AW669666	113161 MA
37	23.4	93.6	438	10	BG364983	BG364983	284002 MA
38	23.4	93.6	487	10	BI454088	BI454088	603174992
39	23.4	93.6	502	10	BI681908	BI681908	164116 MA
40	23.4	93.6	541	10	BE89802	BE89802	187691 MA
41	23.4	93.6	558	9	BB709265	BB709265	BB709265
42	23.4	93.6	561	9	AV612846	AV612846	AV612846
43	23.4	93.6	573	9	AV597535	AV597535	AV597535
44	21.4	85.6	222	10	BE477585	BE477585	161361 BA
45	21.4	85.6	220	9	AA296961	AA296961	EST112514

ALIGNMENTS

RESULT 1

BB666889 BB666889 RIKEN full-length enriched, 260 bp mRNA linear EST 26-OCT-2001
LOCUS Ovary Mus musculus cDNA clone E330035012 5', mRNA sequence.

ACCESSION BB666889
VERSION BB666889
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 260)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyata,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1..260
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E330035012"
/clone.lib="RIKEN full-length enriched, 2 days pregnant adult female ovary"
/sex="female"
/tissue_type="ovary"
/dev_stage="2 days pregnant adult"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATTAATTAATCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 33 a 81 c 88 g 58 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 260;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 187 TCGGATTGCCAGCAATGCGAAGT 211

RESULT

BB856691 330 bp mRNA linear EST 26-NOV-2001
LOCUS BB856691 RIKEN full-length enriched, B16 F10Y cells Mus musculus
DEFINITION cDNA clone G370036K07 5', mRNA sequence.
ACCESSION BB856691
VERSION BB856691.1 GI:17098145
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 330)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,

Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>

M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1..330
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G370036K07"
/clone.lib="RIKEN full-length enriched, B16 F10Y cells"
/cell_type="B16 F10Y cells"

BASE COUNT 48 a 91 c 127 g 64 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 330;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 202 TCGGATTGCCAGCAATGCGAAGT 226

RESULT

AA048794 382 bp mRNA linear EST 09-SEP-1996
LOCUS AA048794 r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:479362 5', mRNA sequence.
ACCESSION AA048794
VERSION AA048794.1 GI:1528466
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 382)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucab,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennan,G., Soares,B., Wilson,R. and

Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:290106
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 344.

FEATURES

source

1..382
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:479362"
/clone_lib="Soares mouse embryo NBMEL3.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5',
TCTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 76 a 102 c 118 g 86 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 382;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 154 TCGATTGCCAGCAATGCGAAGT 178

RESULT 4
AA002532
LOCUS
DEFINITION
AA002532 mouse embryo NBMEL3.5 14.5 Mus musculus cDNA
clone IMAGE:427563 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 406)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE
JOURNAL
COMMENT
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:262115
Seq primer: mob.REGA+ET
High quality sequence stop: 328.

FEATURES

source

1..406
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:427563"
/clone_lib="Soares mouse embryo NBMEL3.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 81 a 107 c 128 g 90 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 406;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 152 TCGATTGCCAGCAATGCGAAGT 176

RESULT 5
AI098985
LOCUS
DEFINITION

AI098985 435 bp mRNA linear EST 20-AUG-1998
uc01g04.r1 Soares mammary_gland_NBMMG Mus musculus cDNA clone
IMAGE:1396758 5' similar to TR:O15412 O15412 CTG4A.; mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.

REFERENCE
AUTHORS
1 (bases 1 to 435)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE
JOURNAL
COMMENT
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Fri May 24 10:49:41 2002

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:908474

Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 329.

Location/Qualifiers

FEATURES

source

1..435
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1396758"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTACCAATCTGAAGTCGAGCGCGCGGAATGTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Patima
Bonaldo."

BASE COUNT 85 a 110 c 134 g 106 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 435;

Best Local Similarity 100.0%; Pred. No. 0.23; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

QY 1 tgcgattgccagcaaatgcgaagt 25
|||||
DB 123 TGGCATTGCCCAACCAATGCGAAGT 147

RESULT

6

BB863087 472 bp mRNA linear EST 27-NOV-2001
LOCUS
BB863087 RIKEN full-length enriched, Leydig cells CRL-2065 MLTC-1
DEFINITION
CDNA Mus musculus cDNA clone G430034D12 5', mRNA sequence.

BB863087

BB863087.1 GI:17109295

EST.

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 472)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii

, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,

Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,

Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,

Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa

, A., Takahashi, F., Takaku-Akaira, S., Tanaka, T., Tomaru, A., Toya, T.

, Wakahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9226

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

e mouse tissues.

FEATURES Location/Qualifiers

1..472

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="G430034D12"

/clone_lib="RIKEN full-length enriched, Leydig cells

CRL-2065 MLTC-1 cDNA"

/cell_type="Leydig cells"

/cell_line="CRL-2065 MLTC-1"

92 a 127 c 156 g 97 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 472;

Best Local Similarity 100.0%; Pred. No. 0.24;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgattgccagcaaatgcgaagt 25

|||||

DB 204 TGGCATTGCCCAACCAATGCGAAGT 228

RESULT

7

BB857917

LOCUS

BB857917 RIKEN full-length enriched, B16 F10Y cells Mus musculus

DEFINITION

CDNA clone G370047K11 5', mRNA sequence.

BB857917

EST.

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 483)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii

, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,

Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,

Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,

Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa

, A., Takahashi, F., Takaku-Akaira, S., Tanaka, T., Tomaru, A., Toya, T.

, Wakahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9222

Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

e mouse tissues.

FEATURES Location/Qualifiers
 source
 1. 483
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370047K11"
 /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 95 a 132 c 159 g 97 t

BASE COUNT 95 a 132 c 159 g 97 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 9; Length 483;
 Best Local Similarity 100.0%; Pred. No. 0.24;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 tgcgattgccagcaaatgcgaagt 25
 |||||
 Db 220 TCGGATTGCCAGCAATGCCGAAGT 244

RESULT 8
 BB857913 485 bp mRNA linear EST 26-NOV-2001
 LOCUS
 DEFINITION BB857913 RIKEN full-length enriched, B16 F10Y cells Mus musculus
 CDNA clone G370047K04 5', mRNA sequence.
 BB857913
 ACCESSION
 VERSION BB857913.1 GI:17099367
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 485)
 Akimura, T., Arawaka, T., Carninci, P., Furuno, M., Hanagaki, T.,
 Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii
 Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
 Salto, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
 Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
 A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
 Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
 2001)

Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

e mouse tissues.

FEATURES Location/Qualifiers
 source
 1. 485
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370047K04"
 /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 101 a 132 c 154 g 98 t

Query Match 100.0%; Score 25; DB 9; Length 485;
 Best Local Similarity 100.0%; Pred. No. 0.24;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 tgcgattgccagcaaatgcgaagt 25
 |||||
 Db 201 TCGGATTGCCAGCAATGCCGAAGT 225

RESULT 9
 BG062828 495 bp mRNA linear EST 25-JAN-2001
 LOCUS
 DEFINITION L0958D12-5 NIA Mouse Newborn Kidney cDNA Library2 (Short) Mus
 musculus CDNA clone L0958D12 5', mRNA sequence.
 BG062828
 ACCESSION
 VERSION BG062828.1 GI:12533696
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 495)
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
 Construction of long-transcript enriched cDNA libraries from
 submicrogram amounts of total RNAs by a universal PCR amplification
 method
 Genome Res. 11 (9), 1553-1558 (2001)
 21429098
 MEDLINE
 COMMENT
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@lgsun.grc.nia.nih.gov
 niaEST (http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html)
 Plate: L0958 row: D column: 12
 Seq primer: -21M13 Reverse
 High quality sequence stop: 495

POLYA-No. Location/Qualifiers
 1. .495
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="niaEST:L0958D12-5"
 /db_xref="taxon:10090"
 /clone="L0958D12"
 /clone_lib="NTA Mouse Newborn Kidney cDNA Library2 (Short)"
 /tissue_type="Newborn Kidney"
 /dev_stage="Newborn"
 /lab_host="DH10B"
 /note=vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a short-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). In brief, double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-PGACTAGTCATGACGACGCGCCGCTTTT-3'] from 26 ug of total RNA, treated with 14 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lona-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex taq polymerase (Takara) with a primer SalI-L. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 1.5 kb. The library was constructed by Yulan Piao(NIA)."
 BASE COUNT 111 a 126 c 152 g 106 t
 ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 495;
 Best Local Similarity 100.0%; Pred. No. 0.24;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tcgcattgccagcaaatgcgaagt 25
 |||||
 Db 138 TGCATTGCCAGCAAAATGCGAAGT 162

RESULT 10
 BB853356
 LOCUS
 DEFINITION BB853356 RIKEN full-length enriched, B16 F10Y cells Mus musculus
 CDNA clone G370012D12 5', mRNA sequence.
 BB853356
 ACCESSION
 VERSION
 KEYWORDS
 EST.
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 502)
 REFERENCE
 AUTHORS
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirose,T., Ito,M., Ito,M., Kawai,J., Kouda,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nomura,K., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,A., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
 TITLE
 JOURNAL
 COMMENT
 Unpublished (2001)
 Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
 e mouse tissues.

FEATURES
 Location/Qualifiers
 1. 502
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370012D12"
 /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 BASE COUNT 103 a 136 c 161 g 102 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 502;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tcgcattgccagcaaatgcgaagt 25
 |||||
 Db 197 TGCATTGCCAGCAAAATGCGAAGT 221

RESULT 11
 BB605812
 LOCUS
 DEFINITION BB605812 RIKEN full-length enriched, 0 day neonate lung Mus musculus CDNA clone E030013D02 5', mRNA sequence.
 BB605812
 ACCESSION
 VERSION
 KEYWORDS
 EST.
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 608)
 REFERENCE
 AUTHORS
 Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 TITLE
 JOURNAL
 COMMENT
 Unpublished (2001)
 On Dec 5, 2000 this sequence version replaced gi:11557214.
 Contact: Yoshihide Hayashizaki

contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGCGCGCCGACTCGAGTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTAATTAATTAATCCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I." 127 a 164 c 205 g 126 t 1 others

BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.27; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 197 TCGGATTGCCAGCAATGCGAGT 221

RESULT 13
BB633565 625 bp mRNA linear EST 26-OCT-2001
LOCUS BB633565 RIKEN full-length enriched, adult male spinal cord Mus
DEFINITION musculus cDNA clone A330062B19 5', mRNA sequence.

ACCESSION BB633565
VERSION BB633565.1 GI:16469989
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Akakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, K., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
waji, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Tameda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamashita, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

TITLE
JOURNAL
COMMENT

Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001).
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES
source

Location/Qualifiers
1..625
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A330062B19"
/clone_lib="RIKEN full-length enriched, adult male spinal cord"
/sex="male"
/tissue_type="spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGATCTCGAGTTAATTAATTAATCCGCCGCCGCC 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTAATTAATTAATCCGCCGCCGCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 145 a 152 c 200 g 128 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 625;
Best Local Similarity 100.0%; Pred. No. 0.27; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 145 TCGGATTGCCAGCAATGCGAGT 169

RESULT 14
BB637089 641 bp mRNA linear EST 26-OCT-2001
LOCUS BB637089 RIKEN full-length enriched, adult male aorta and vein Mus
DEFINITION musculus cDNA clone A530052E13 5', mRNA sequence.

ACCESSION BB637089
VERSION BB637089.1 GI:16472874
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Akakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)

TITLE
JOURNAL
COMMENT

transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTCCTCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT
ORIGIN

138 a 162 c 216 g 127 t

Query Match 100.0%; Score 25; DB 9; Length 643;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcattgccagcaaatgcgaagt 25
|||||
Db 209 TCGATTGCCAGCAAAATGCGAAGT 233

Search completed: May 23, 2002, 23:05:31
Job time: 7487 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 24, 2002, 00:18:05 ; Search time 121.39 Seconds
(without alignments)
50.588 Million cell updates/sec

Title: US-09-696-686-46
Perfect score: 25
Sequence: 1 tgcgattgccagcaaatgcgaagt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.4	85.6	744	4	US-09-163-285-3
2	21.4	85.6	1512	4	US-09-163-285-1
3	16.4	65.6	4195	1	US-08-200-016-1
4	16.2	64.8	1980	1	US-08-350-741-1
5	16.2	64.8	1980	2	US-08-463-875A-1
6	16	64.0	361	2	US-08-465-380-39
7	16	64.0	361	2	US-08-486-397-39
8	16	64.0	361	2	US-08-486-399-39
9	16	64.0	361	2	US-08-461-365-39
10	16	64.0	361	2	US-08-634-641-39
11	16	64.0	361	3	US-09-249-471-39
12	16	64.0	361	3	US-09-249-472-39
13	16	64.0	361	3	US-09-249-451-39
14	16	64.0	361	3	US-08-809-455-39
15	16	64.0	361	3	US-09-249-461-39
16	16	64.0	361	3	US-09-249-448-39
17	16	64.0	2941	4	US-09-425-383-1
18	16	64.0	4411529	4	US-09-103-840A-1
19	15.8	63.2	1080	3	US-09-150-133-12
20	15.8	63.2	1080	3	US-09-150-141-12
21	15.8	63.2	1080	4	US-09-374-493-12
22	15.8	63.2	1080	4	US-09-374-824-12
23	15.8	63.2	1080	4	US-09-374-492-12
24	15.6	62.4	1854	3	US-08-923-454A-13
25	15.6	62.4	4928	1	US-08-399-561-1
26	15.4	61.6	1160	2	US-08-955-138-5
27	15.4	61.6	1835	4	US-09-338-671-3

28	15.4	61.6	9542	4	US-08-968-685A-9	Sequence 9, Appli
29	15.4	61.6	36741	4	US-09-301-865-3	Sequence 3, Appli
30	15.2	60.8	2391	2	US-08-817-900-1	Sequence 1, Appli
31	15.2	60.8	2391	4	US-09-236-645-1	Sequence 1, Appli
32	15.2	60.8	3070	1	US-08-428-732-3	Sequence 3, Appli
33	15.2	60.8	3147	4	US-09-153-804-12	Sequence 12, Appli
34	15.2	60.8	3180	1	US-08-480-662-1	Sequence 1, Appli
35	15.2	60.8	3180	3	US-08-918-190-1	Sequence 1, Appli
36	15.2	60.8	3180	4	US-09-234-232-1	Sequence 1, Appli
37	15.2	60.8	3180	5	PCT-US96-09927-1	Sequence 1, Appli
38	15	60.0	993	2	US-08-245-511-57	Sequence 57, Appli
39	15	60.0	994	2	US-08-600-993A-57	Sequence 57, Appli
40	15	60.0	2464	1	US-07-863-169A-4	Sequence 4, Appli
41	15	60.0	2464	2	US-08-429-964-4	Sequence 4, Appli
42	15	60.0	2464	3	US-07-935-087-4	Sequence 4, Appli
43	15	60.0	2464	5	PCT-US93-08062-4	Sequence 4, Appli
44	15	60.0	3180	1	US-07-944-525-1	Sequence 1, Appli
45	15	60.0	4011	1	US-08-121-057-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-163-285-3
; Sequence 3, Application US/09163285
; Patent No. 6204013
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/090,398
; FILING DATE: June 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..744
; US-09-163-285-3

Query Match 85.6%; Score 21.4; DB 4; Length 744;
Best Local Similarity 95.7%; Pred. No. 0.086;

;; FILING DATE: 09-MAY-1994
;; APPLICATION NUMBER: US 07/952,737
;; FILING DATE: 30-NOV-1992
;; APPLICATION NUMBER: GB 9007194.5
;; FILING DATE: 30-MAR-1990
;; APPLICATION NUMBER: PCT/GB91/00484
;; FILING DATE: 28-MAR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: WILSON, MARY J.
;; REGISTRATION NUMBER: 32,955
;; REFERENCE/DOCKET NUMBER: 117-158
;; TELEPHONE: (703) 816-4000
;; TELEFAX: (703) 816-4100
;; TELEX: 200797 NIXN UR
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1980 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 395..1822
US-08-350-741-1

Query Match 64.8%; Score 16.2; DB 1; Length 1980;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 gattgccagcaaatgcgaag 24
|| ||||| |||| |||| ||||
Db 496 GACTGCCCAGCAGATGCCAAG 516

RESULT 5
US-08-463-875A-1
; Sequence 1, Application US/08463875A
; Patent No. 5980907
; GENERAL INFORMATION:
; APPLICANT: DOUGAN, Gordon
; APPLICANT: CHARLES, Ian G.
; APPLICANT: HORMACHE, Carlos E.
; APPLICANT: JOHNSON, Kevin S.
; APPLICANT: CHATFIELD, Steven N.
; TITLE OF INVENTION: LIVE VACCINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON and VANDERHVE PC
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,875A
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/340,741
; FILING DATE: 07-DEC-1994
; APPLICATION NUMBER: US 07/952,737
; FILING DATE: 30-NOV-1992
; APPLICATION NUMBER: GB 9007194.5
; FILING DATE: 30-MAR-1990
; APPLICATION NUMBER: PCT/GB91/00484
; FILING DATE: 28-MAR-1991

;; ATTORNEY/AGENT INFORMATION:
;; NAME: WILSON, MARY J.
;; REGISTRATION NUMBER: 32,955
;; REFERENCE/DOCKET NUMBER: 117-158
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 816-4000
;; TELEFAX: (703) 816-4100
;; TELEX: 200797 NIXN UR
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1980 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 395..1822
US-08-463-875A-1

Query Match 64.8%; Score 16.2; DB 2; Length 1980;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 gattgccagcaaatgcgaag 24
|| ||||| |||| |||| ||||
Db 496 GACTGCCCAGCAGATGCCAAG 516

RESULT 6
US-08-465-380-39
; Sequence 39, Application US/08465380
; Patent No. 5863894
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter W. Bergum
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,380
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 213/268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:

us-09-696-686-46.rni

Fri May 24 10:49:40 2002

```

;
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
; US-08-486-397-39

Query Match 64.0%; Score 16; DB 2; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gcgattgcccagcaaatgcgaagt 25
| | | | | | | | | | | | | | |
Db 74 GAGATTGTCCAGCAAAATGAGGAAT 97

RESULT 8
US-08-486-399-39
; Sequence 39, Application US/08486399
; Patent No. 5866543
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter W. Bergum
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,399
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 213/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEX: (213) 955-0440
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
; US-08-486-399-39

Query Match 64.0%; Score 16; DB 2; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gcgattgcccagcaaatgcgaagt 25
| | | | | | | | | | | | | | |
Db 74 GAGATTGTCCAGCAAAATGAGGAAT 97

RESULT 7
US-08-486-397-39
; Sequence 39, Application US/08486397
; Patent No. 5866542
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter W. Bergum
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; NUMBER OF SEQUENCES: 357
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,397
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 213/269
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEX: (213) 955-0440
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
```


APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORGANISM: Necator americanus
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 16...252
US-08-809-455-39

Query Match 64.0%; Score 16; DB 3; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gcgattgccagcaaatcggaagt 25
| | | | | | | | | | | | | | | | | | | | | |
DB 74 GAGATTGCCAGCAAAATGAGGAAT 97

RESULT 15
US-09-249-461-39
Sequence 39, Application US/09249461
Patent No. 6096877
GENERAL INFORMATION:
APPLICANT: Vlasuk, George Phillip
APPLICANT: Stanssens, Patrick Eric Hugo
APPLICANT: Messens, Joris Hilda Lieven
APPLICANT: Lauwereys, Marc Josef
APPLICANT: Laroche, Yves Rene
APPLICANT: Jespers, Laurent Stephane
APPLICANT: Gansemans, Yannick Georges Jozef
APPLICANT: Moyle, Matthew
APPLICANT: Berquim, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
TITLE OF INVENTION: INHIBITORS AND ANTICOGULANT
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,461
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,455
FILING DATE: April 17, 1997
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995

APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/465,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: Necator americanus
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 16...252
US-09-249-461-39

Query Match 64.0%; Score 16; DB 3; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gcgattgccagcaaatcggaagt 25
| | | | | | | | | | | | | | | | | | | | | |
DB 74 GAGATTGCCAGCAAAATGAGGAAT 97

Search completed: May 24, 2002, 00:18:10
Job time: 11556 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:23:09 ; Search time 525.6 seconds
(without alignments)
81.664 Million cell updates/sec

Title: US-09-696-686-46

Perfect score: 25

Sequence: 1 tgcgattgccagcaaatgcgaagt 25

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	22	AA05281
2	25	100.0	25	22	AA05288
3	25	100.0	49	22	AA05283
4	25	100.0	1839	22	AA05278
5	25	100.0	1848	22	AA05286
6	24	96.0	370	22	AA05285
7	21.4	85.6	238	21	AA04529
8	21.4	85.6	371	20	AA08062
9	21.4	85.6	443	20	AA08063

10	21.4	85.6	510	22	AA093625
11	21.4	85.6	744	21	AA049678
12	21.4	85.6	746	22	AA041104
13	21.4	85.6	747	22	AA090704
14	21.4	85.6	1306	20	AA024902
15	21.4	85.6	1312	22	AA050462
16	21.4	85.6	1315	22	AA050464
17	21.4	85.6	1419	22	AA093914
18	21.4	85.6	1457	20	AA078923
19	21.4	85.6	1462	21	AA016621
20	21.4	85.6	1470	21	AA096349
21	21.4	85.6	1470	21	AA058116
22	21.4	85.6	1470	22	AA046164
23	21.4	85.6	1512	21	AA049677
24	21.4	85.6	1623	20	AA088064
25	20.2	80.8	1362	22	AA050279
26	20.2	80.8	1675	22	AA021311
27	20.2	80.8	10140	22	AA049201
28	19.2	76.8	1657	22	AA040377
29	18.8	75.2	568	22	AA013480
30	18.8	75.2	2096	22	AA010871
31	18.8	75.2	2096	22	AA010871
32	18.2	72.8	107	21	AA023001
33	18.2	72.8	541	22	AA092573
34	18.2	72.8	1285	22	ABL17251
35	18.2	72.8	2678	22	AA04725
36	18.2	72.8	4321	23	ABL17250
37	18.2	72.8	4487	21	AA076070
38	17.2	68.8	3154	23	ABL18228
39	17.2	68.8	18748	23	ABL05146
40	17	68.0	152	23	AA050725
41	17	68.0	377	18	AA078311
42	17	68.0	424	18	AA078020
43	17	68.0	1071	23	AA051969
44	17	68.0	1080	23	AA054722
45	17	68.0	3899	23	ABL15747

ALIGNMENTS

RESULT 1
AA05281
ID AA05281 standard; DNA; 25 BP.
XX
AC AA05281;
XX
DT 07-SEP-2001 (first entry)
XX
DE PCR primer #2 used to amplify T243 homologous sequences.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; PCR primer; ss.
XX Mus sp.
XX OS
XX WO200130798-A1.
XX
XX
PD 03-MAY-2001.
XX
XX
PF 26-OCT-2000; 2000WO-US29382.
XX
XX
PR 26-OCT-1999; 99US-0161488.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Klein R, Matthews W, Moore M, Allen KD;
XX
XX WPI; 2001-300473/31.
XX
XX Novel transgenic animals useful as animal model for characterization of
PT function of a gene encoding trinucleotide repeat proteins (TRPs).

Umbilical vein end
Human myocardium s
Primer specific fo
Human secretory pr
Human secreted pro
Human secreted pro
Human secreted pro
Human cDNA encodin
Human zcalc-1 DNA
Human secreted pro
cDNA encoding a no
Human PR04354 nucl
Human DNA encoding
Human myocardium s
Human zcalc-1 DNA
Human trinucleotid
Human cDNA sequenc
Human CHOT promote
Human secreted pro
Human cDNA clone (c
Human cDNA sequenc
Human genomic DNA
Human secreted pro
Human cDNA 3'-end
Drosophila melanog
Human full-length
Drosophila melanog
Human ORFX ORF1625
Drosophila melanog
Drosophila melanog
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Drosophila melanog

PT contains heterozygous disruption in a gene encoding TRP -
 PS Claim 14; Fig 13; 106pp; English.
 XX The present sequence for PCR primer #2 is used to amplify sequences
 CC homologous to the T243 gene which encodes for a trinucleotide repeat
 CC protein (TRP). The invention describes methods of producing embryonic
 CC stem (ES) cells comprising a heterozygous disruption in a target DNA
 CC sequence (preferably gene T243) encoding a TRP and of producing a
 CC knockout mouse comprising a homozygous disruption in a gene encoding TRP,
 CC where the disruption inhibits the production of the wild type TRP. The
 CC invention also relates to identifying agents capable of affecting a
 CC phenotype of a knockout mouse. Also described are methods of determining
 CC whether expansion of the trinucleotide repeat in a gene encoding TRP
 CC produces a phenotypic change. The transgenic animals and the cells are
 CC useful for identifying compounds capable of ameliorating disease
 CC symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.

XX SQ Sequence 25 BP; 7 A; 6 C; 7 G; 5 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.0048;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
 |||||
 Db 1 tgcgattgccagcaaatgcgaagt 25

RESULT 2
 AAS05288/c
 ID AAS05288 standard; DNA; 25 BP.
 XX
 AC AAS05288;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE Mouse ES cell T243 gene PCR primer #432.
 XX
 KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW mouse; PCR primer; ss.
 XX
 OS Mus sp.
 XX
 PN WO200130798-A1.
 XX
 PD 03-MAY-2001.
 XX
 PF 26-OCT-2000; 2000WO-US29382.
 XX
 PR 26-OCT-1999; 99US-0161488.
 XX
 PR 26-OCT-1999; 99US-0161488.
 XX
 PA (DELT-) DELTAGEN INC.
 XX
 PI Klein R, Matthews W, Moore M, Allen KD;
 XX
 DR WPT; 2001-300473/31.
 XX
 XX Novel transgenic animals useful as animal model for characterization of
 XX function of a gene encoding trinucleotide repeat proteins (TRPs),
 XX contains heterozygous disruption in a gene encoding TRP -
 PS Example 12; Fig 13; 106pp; English.

XX The present sequence for PCR primer #432 is used to amplify the T243
 CC gene from an F1 mouse embryonic stem cell genomic library. The T243 gene
 CC encodes for a trinucleotide repeat protein (TRP). The invention describes
 CC methods of producing embryonic stem (ES) cells comprising a heterozygous
 CC disruption in a target DNA sequence (preferably T243) encoding a TRP and
 CC of producing a knockout mouse comprising a homozygous disruption in a
 CC gene encoding TRP, where the disruption inhibits the production of the
 CC wild type TRP. The invention also relates to identifying agents capable
 CC of affecting a phenotype of a knockout mouse. Also described are methods
 CC of determining whether expansion of the trinucleotide repeat in a gene
 CC encoding TRP produces a phenotypic change. The transgenic animals and
 CC the cells are useful for identifying compounds capable of ameliorating
 CC disease symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.

XX SQ Sequence 25 BP; 5 A; 7 C; 6 G; 7 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.0048;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
 |||||
 Db 25 TGCgATTGCCcAGCAAAATGCGAAGT 1

RESULT 3
 AAS05283
 ID AAS05283 standard; DNA; 49 BP.
 XX
 AC AAS05283;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE PCR primer #489 used to amplify T243 PCR product from pool A7.
 XX
 KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; mouse; PCR primer; ss.
 XX
 OS Mus sp.
 XX
 PN WO200130798-A1.
 XX
 PD 03-MAY-2001.
 XX
 PF 26-OCT-2000; 2000WO-US29382.
 XX
 PR 26-OCT-1999; 99US-0161488.
 XX
 PA (DELT-) DELTAGEN INC.
 XX
 PI Klein R, Matthews W, Moore M, Allen KD;
 XX
 DR WPT; 2001-300473/31.
 XX
 XX Novel transgenic animals useful as animal model for characterization of
 XX function of a gene encoding trinucleotide repeat proteins (TRPs),
 XX contains heterozygous disruption in a gene encoding TRP -
 PS Example 12; Fig 13; 106pp; English.
 XX The present sequence for PCR primer #489 is used to further amplify
 CC a T243 gene PCR product from pool A7 (obtained by PCR of the T243 gene

CC from an R1 mouse embryonic stem cell genomic library). Gene T243 encodes
CC for a trinucleotide repeat protein (TRP). The invention describes methods
CC of producing embryonic stem (ES) cells comprising a heterozygous
CC disruption in a target DNA sequence (preferably gene T243) encoding a
CC TRP and of producing a knockout mouse comprising a homozygous disruption
CC in a gene encoding TRP, where the disruption inhibits the production of
CC the wild type TRP. The invention also relates to identifying agents
CC capable of affecting a phenotype of a knockout mouse. The transgenic
CC animals and the cells are useful for identifying compounds capable of
CC ameliorating disease symptoms, and as test substrates for the
CC identification of drugs, pharmaceuticals, therapies and interventions
CC which may be effective in treating trinucleotide repeat disorders e.g.
CC fragile X syndrome and Huntington's disease. The animal models for
CC trinucleotide repeat disorders are ideal model systems to study the
CC progression of disease in vivo, the molecular basis of these diseases
CC and show the features observed in human disease. Using the mice, it is
CC possible to model both the pathogenic mechanism and the trinucleotide
CC repeat instability in the mouse.
XX
SQ Sequence 49 BP; 9 A; 13 C; 14 G; 13 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 49;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgcccgacaaatgcgaagt 25
Db 25 tgcgattgcccgacaaatgcgaagt 49
|||||

RESULT 4
AAS05278
ID AAS05278 standard; cDNA; 1839 BP.
XX
AC AAS05278;
XX
DT 07-SEP-2001 (first entry)
XX
DE Murine trinucleotide repeat protein (TRP) cDNA sequence.
XX
KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; murine; ss.
XX
OS Mus sp.

XX
FH Key Location/Qualifiers
FT CDS 55..885
FT /*tag= a
FT /product= "TRP"

XX
PN WO200130798-A1.
XX
XX 03-MAY-2001.
XX
XX 26-OCT-2000; 2000WO-US29382.
XX
XX 26-OCT-1999; 99US-0161488.
XX
XX (DELTA-) DELTAGEN INC.
XX
XX Klein R, Matthews W, Moore M, Allen KD;
XX
XX WPI; 2001-300473/31.
DR P-PSDB; AAU02498.
XX
XX Novel transgenic animals useful as animal model for characterization of
XX function of a gene encoding trinucleotide repeat proteins (TRPs),
XX contains heterozygous disruption in a gene encoding TRP -
XX
XX Claim 8; Fig 11; 106pp; English.

CC The present sequence encodes for murine trinucleotide repeat protein
CC (TRP). The invention describes methods of producing embryonic stem (ES)
CC cells comprising a heterozygous disruption in a target DNA sequence
CC (preferably gene T243) encoding a TRP and of producing a knockout mouse
CC comprising a homozygous disruption in a gene encoding TRP, where the
CC disruption inhibits the production of the wild type TRP. The invention
CC also relates to identifying agents capable of affecting a phenotype of
CC a knockout mouse. Also described are methods of determining whether
CC expansion of the trinucleotide repeat in a gene encoding TRP produces a
CC phenotypic change. The transgenic animals and the cells are useful for
CC identifying compounds capable of ameliorating disease symptoms, and as
CC test substrates for the identification of drugs, pharmaceuticals,
CC therapies and interventions which may be effective in treating
CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
CC disease. The animal models for trinucleotide repeat disorders are ideal
CC model systems to study the progression of disease in vivo, the molecular
CC basis of these diseases and show the features observed in human disease.
CC Using the mice, it is possible to model both the pathogenic mechanism and
CC the trinucleotide repeat instability in the mouse.
XX
SQ Sequence 1839 BP; 424 A; 522 C; 535 G; 358 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 1839;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgcccgacaaatgcgaagt 25
Db 182 tgcgattgcccgacaaatgcgaagt 206
|||||

RESULT 5
AAS05286
ID AAS05286 standard; DNA; 1848 BP.
XX
AC AAS05286;
XX
DT 07-SEP-2001 (first entry)
XX
DE DNA sequence of an expanded T243 gene.
XX
DE Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; mutant; ds.
XX
OS Mus sp.
OS Synthetic.

XX
FH Key Location/Qualifiers
FT CDS 55..894
FT /*tag= a
FT /product= "Expanded TRP"

XX
PN WO200130798-A1.
XX
XX 03-MAY-2001.
XX
XX 26-OCT-2000; 2000WO-US29382.
XX
XX 26-OCT-1999; 99US-0161488.
XX
XX (DELTA-) DELTAGEN INC.
XX
XX Klein R, Matthews W, Moore M, Allen KD;
XX
XX WPI; 2001-300473/31.
DR P-PSDB; AAU02500.

XX
XX Novel transgenic animals useful as animal model for characterization of
XX function of a gene encoding trinucleotide repeat proteins (TRPs),
XX contains heterozygous disruption in a gene encoding TRP -
XX
XX

PS Disclosure; Fig 15; 106pp; English.

XX The present sequence represents the DNA sequence of an expanded T243 gene. The T243 gene encodes for a trinucleotide repeat protein (TRP).

CC The invention describes methods of producing embryonic stem (ES) cells comprising a heterozygous disruption in a target DNA sequence (preferably gene T243) encoding a TRP and of producing a knockout mouse comprising a heterozygous disruption in a gene encoding TRP, where the disruption inhibits the production of the wild type TRP. The invention also relates to identifying agents capable of affecting a phenotype of a knockout mouse. Also described are methods of determining whether expansion of the trinucleotide repeat in a gene encoding TRP produces a phenotypic change. The transgenic animals and the cells are useful for identifying compounds capable of ameliorating disease symptoms, and as test substrates for the identification of drugs, pharmaceuticals, and therapies and interventions which may be effective in treating trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's disease. The animal models for trinucleotide repeat disorders are ideal model systems to study the progression of disease in vivo, the molecular basis of these diseases and show the features observed in human disease. Using the mice, it is possible to model both the pathogenic mechanism and the trinucleotide repeat instability in the mouse.

XX Sequence 1848 BP; 415 A; 519 C; 543 G; 371 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 1848;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 191 tgcgattgccagcaaatgcgaagt 215

RESULT 6
AAS05285
ID AAS05285 standard; DNA; 370 BP.

XX AAS05285;

XX 07-SEP-2001 (first entry)

DE T243 gene homologous sequence #2 generated by PCR.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;

XX transgenic animal; knockout mouse; triplet repeat expansion;

XX fragile X syndrome; Huntington's disease; mouse; ds.

XX Mus sp.

XX WO200130798-A1.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-US29382.

XX 26-OCT-1999; 99US-0161488.

XX (DELT-) DELTAGEN INC.

XX Klein R, Matthews W, Moore M, Allen KD;

XX WPI; 2001-300473/31.

XX Novel transgenic animals useful as animal model for characterization of function of a gene encoding trinucleotide repeat proteins (TRPs),
XX contains heterozygous disruption in a gene encoding TRP -

XX Claim 11; Fig 14; 106pp; English.

XX The present sequence for T243 homologous sequence #2 is generated by PCR and is homologous to the T243 gene which encodes for a trinucleotide

CC repeat protein (TRP). The invention describes methods of producing embryonic stem (ES) cells comprising a heterozygous disruption in a target DNA sequence (preferably gene T243) encoding a TRP and of producing a knockout mouse comprising a heterozygous disruption in a gene encoding TRP, where the disruption inhibits the production of the wild type TRP. The invention also relates to identifying agents capable of affecting a phenotype of a knockout mouse. Also described are methods of determining whether expansion of the trinucleotide repeat in a gene encoding TRP produces a phenotypic change. The transgenic animals and the cells are useful for identifying compounds capable of ameliorating disease symptoms, and as test substrates for the identification of drugs, pharmaceuticals, therapies and interventions which may be effective in treating trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's disease. The animal models for trinucleotide repeat disorders are ideal model systems to study the progression of disease in vivo, the molecular basis of these diseases and show the features observed in human disease. Using the mice, it is possible to model both the pathogenic mechanism and the trinucleotide repeat instability in the mouse.

XX Sequence 370 BP; 73 A; 86 C; 149 G; 62 T; 0 other;

Query Match 96.0%; Score 24; DB 22; Length 370;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaag 24
|||||
Db 1 tgcgattgccagcaaatgcgaag 24

RESULT 7
AAA45529

ID AAA45529 standard; cDNA; 238 BP.

XX AAA45529;

XX 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:2104.

XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
XX antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24206.

XX 15-OCT-1998; 98US-0104436.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Bowman MR;

XX WPI; 2000-317938/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 729; 803pp; English.
XX
CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (SESTs). Isolated from human, mouse, chicken and rat
CC tissue sources, the SESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytotstatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnery; antitumor; osteoporotic; neuroprotective;
CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidiabetic. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 238 BP; 51 A; 51 C; 80 G; 56 T; 0 other;

Query Match 85.6%; Score 21.4; DB 21; Length 238;
Best Local Similarity 95.7%; Pred. No. 0.46;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 cgattgcccagcaaatgcgaagt 25
|| |||||
DB 155 cgcttgcgcagcaaatgcgaagt 177

RESULT 8
AAAX8062
ID AAAX8062 standard; DNA; 371 BP.
XX
AC AAAX8062;
XX
DT 09-SEP-1999 (first entry)
XX
DE Human zcalc-1 DNA allele 1.
XX
KW Calcitonin-like protein 1; zcalc-1; human; osteopathic; antidiabetic;
KW hypotensive; vasodilator; neuromodulator; peripheral organ; treatment;
KW Paget's disease; hypercalcaemia; osteoporosis; Raynaud's disease;
KW type I diabetes; hypertension; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..371
FT /*tag= a
FT /product= "zcalc-1"
FT /note= "calcitonin-like protein 1"
XX
PN WO9931131-A2.
XX
XX 24-JUN-1999.
XX
PF 18-DEC-1998; 98WO-US26940.
XX
PR 18-DEC-1997; 97US-0993935.
XX

XX (ZYMO) ZYMOGENETICS INC.
PA Moore EE, Raymond FC, Sheppard PO;
PI WPI; 1999-418753/35.
XX P-PSDB; AAY25418.
DR
DR New mammalian nucleic acid encoding a calcitonin-like polypeptide-1
XX
PS Disclosure; Page 47-48; 55pp; English.
XX
CC This invention describes a novel human polynucleotide that encodes a
CC calcitonin-like polypeptide which has osteopathic, antidiabetic and
CC hypotensive activity and also acts as a vasodilator. Antibodies derived
CC from the product of the invention can be used to isolate and purify
CC the polypeptide sequences. The polypeptides can be used as
CC neuromodulators in a variety of peripheral organs. They can be used in
CC the treatment of Paget's disease, hypercalcaemia, and osteoporosis. It
CC can also be used to treat Raynaud's disease, and may be able to inhibit
CC the progression of type I diabetes. The polypeptide sequences may be
CC administered as vasodilators to treat hypertension.
XX
SQ Sequence 371 BP; 79 A; 81 C; 125 G; 86 T; 0 other;

Query Match 85.6%; Score 21.4; DB 20; Length 371;
Best Local Similarity 95.7%; Pred. No. 0.5;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 cgattgcccagcaaatgcgaagt 25
|| |||||
DB 222 cgcttgcgcagcaaatgcgaagt 244
RESULT 9
AAAX8063
ID AAAX8063 standard; DNA; 443 BP.
XX
AC AAAX8063;
XX
DT 09-SEP-1999 (first entry)
XX
DE Human zcalc-1 DNA allele 2.
XX
KW Calcitonin-like protein 1; zcalc-1; human; osteopathic; antidiabetic;
KW hypotensive; vasodilator; neuromodulator; peripheral organ; treatment;
KW Paget's disease; hypercalcaemia; osteoporosis; Raynaud's disease;
KW type I diabetes; hypertension; ss.
XX
OS Homo sapiens.
XX
PN WO9931131-A2.
XX
XX 24-JUN-1999.
XX
PF 18-DEC-1998; 98WO-US26940.
XX
PR 18-DEC-1997; 97US-0993935.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Moore EE, Raymond FC, Sheppard PO;
XX
XX WPI; 1999-418753/35.
DR
DR New mammalian nucleic acid encoding a calcitonin-like polypeptide-1
XX
PS Disclosure; Page 48; 55pp; English.
XX
CC This invention describes a novel human polynucleotide that encodes a
CC calcitonin-like polypeptide which has osteopathic, antidiabetic and
CC hypotensive activity and also acts as a vasodilator. Antibodies derived

CC from the product of the invention can be used to isolate and purify
CC the polypeptide sequences. The polypeptides can be used as
CC neuromodulators in a variety of peripheral organs. They can be used in
CC the treatment of Paget's disease, hypercalcaemia, and osteoporosis. They
CC can also be used to treat Raynaud's disease, and may be able to inhibit
CC the progression of type I diabetes. The polypeptide sequences may be
CC administered as vasodilators to treat hypertension.

XX Sequence 443 BP; 113 A; 94 C; 142 G; 94 T; 0 other;

Query Match 85.6%; Score 21.4; DB 20; Length 443;
Best Local Similarity 95.7%; Pred. No. 0.52;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaagt 25
|| |||||

DB 182 cgcttgcccagcaaatgcgaagt 204
|| |||||

RESULT 10
AAAF93625
ID AAF93625 standard; cDNA; 510 BP.

XX AAF93625;
XX 21-MAY-2001 (first entry)

XX Umbilical vein endothelial cell cDNA encoding SRT protein SEQ ID 446.
DE Human; SRT; gene therapy; gene mapping; tissue typing; ss.

KW Homo sapiens.

OS WO200107611-A2.

XX 01-FEB-2001.

XX 21-JUL-2000; 2000WO-US20006.

XX 26-JUL-1999; 99US-0145701.

XX (GETH) GENENTECH INC.

XX Baker KP, Goddard A, Wood WI;

XX WPI; 2001-112729/12.

XX New isolated nucleic acid molecule encoding a SRT polypeptide is useful
PT for production of recombinant SRT polypeptides, gene mapping,
PT diagnosing genetic disorders and for gene therapy -

XX Claim 2; Fig 446; 663pp; English.

XX Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
CC human SRT proteins. The cDNA sequences are isolated from various
CC different human tissue cDNA libraries. The invention relates to a method
CC for detecting cDNA encoding an SRT protein, a vector containing cDNA
CC encoding SRT, a host cell transformed with the vector, an isolated SRT
CC polypeptide, and an antibody which binds to SRT. The polynucleotide
CC sequence can be used in gene therapy and is useful in the recombinant
CC production of SRT polypeptides, as a hybridisation probe to screen
CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
CC map the gene encoding the SRT polypeptides and analysing genetic
CC disorders, tissue typing and disease tissue detection. The SRT
CC polynucleotide sequences can be used in polymerase chain reaction,
CC screening for new therapeutic molecules and generation of antisense RNA
CC and DNA.

XX Sequence 510 BP; 122 A; 114 C; 168 G; 104 T; 2 other;

Query Match 85.6%; Score 21.4; DB 22; Length 510;

Best Local Similarity 95.7%; Pred. No. 0.53;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaagt 25
|| |||||

DB 206 cgcttgcccagcaaatgcgaagt 228
|| |||||

RESULT 11
AAZ49678
ID AAZ49678 standard; cDNA; 744 BP.

XX AAZ49678;

XX 07-APR-2000 (first entry)

XX Human myocardium secreted protein-5 coding region.

XX Human; myocardium secreted protein-5; MSP-5; hypotensive; cardiant;
KW cardiac cellular process; cardiovascular disorder; cardiomyopathy;
KW congestive heart failure; hypertension; atherosclerosis;
KW coronary artery disease; valvular disease; arrhythmia; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FH 1..744
FT mat_peptide

FT /*tag= a /product= "MSP-5"

FT /*note= "Myocardium secreted protein-5"

FT sig_peptide 1..66

FT /*tag= b 67..744

FT mat_peptide /*tag= c

FT /*label= Mature_MSP-5

XX W09967385-A1.

XX 29-DEC-1999.

XX 22-JUN-1999; 99WO-US13937.

XX 23-JUN-1998; 98US-0090398.

XX 29-SEP-1998; 98US-0163285.

XX (MILL-) MILLENNIUM PHARM INC.

XX Khodadoust MM;

XX WPI; 2000-136983/12.

XX P-PSDB; AAY44629.

XX Novel myocardium secreted protein-5 polynucleotides, used to modulate a
PT variety of cellular processes -

XX Claim 1; Pages 92-93; 99pp; English.

XX The present sequence is the coding region of myocardium secreted protein
PS -5 (MSP-5) cDNA which was isolated from a cDNA library prepared from a
CC cardiac tissue sample obtained from a biopsy of a 42 year old woman
CC suffering from congestive heart failure. The MSP-5 protein has
CC hypertensive and cardiant activities. It is highly expressed in heart,
CC brain, placenta, fetal lung, liver, kidney, testis, small intestine and
CC pituitary gland. The present sequence is used to modulate a variety of
CC cellular processes, especially cardiac cellular processes. MSP-5 is used
CC to modulate the activity of one or more proteins involved in a
CC cardiovascular disorder, e.g. congestive heart failure or cardiomyopathy.
CC Conditions and diseases which can be treated include hypertension,
CC atherosclerosis, coronary artery spasm, coronary artery disease, valvular
CC disease, arrhythmias, and cardiopathies (e.g. hypertrophic, dilative, or
CC restrictive cardiomyopathies), and disorders related to under or over
CC expression of MSP-5.

XX

RESULT 14

AAZ24902
ID AA224902 standard; DNA; 1306 BP.
XX
XX
AC AAZ24902;
XX
DT 02-DEC-1999 (first entry)
XX
DE Human secreted protein gene 92 clone HATAC53.
XX
XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
XX WO9947540-A1.
XX
XX 23-SEP-1999.
XX
XX 18-MAR-1999; 99WO-US05804.
XX
XX 19-MAR-1998; 98US-0078563.
XX 19-MAR-1998; 98US-0078566.
XX 19-MAR-1998; 98US-0078573.
XX 19-MAR-1998; 98US-0078574.
XX 19-MAR-1998; 98US-0078576.
XX 19-MAR-1998; 98US-0078577.
XX 19-MAR-1998; 98US-0078578.
XX 19-MAR-1998; 98US-0078579.
XX 19-MAR-1998; 98US-0078581.
XX 01-APR-1998; 98US-0080312.
XX 01-APR-1998; 98US-0080313.
XX 01-APR-1998; 98US-0080314.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW;
PI Olsen HS, Shi Y, Moore PA;
XX
XX WPI; 1999-562050/47.
XX P-PSDB; AAY41399.
XX
XX New isolated human genes, useful for diagnosis and treatment of e.g.
PT cancers, neurological disorders, immune diseases, inflammation or blood
PT disorders -
XX
XX Claim 1; Page 353; 484pp; English.
XX
XX This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number, and the clone it is derived
CC from, are detailed in the descriptor line. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. AA224802) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 95 novel genes and their fragments (nucleic
CC acid sequences: AA224811-224907; amino acid sequences AAY41308-Y41404)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 95
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AA224811 for described uses).
XX
XX Sequence 1306 BP; 334 A; 291 C; 377 G; 300 T; 4 other;

Query Match 85.6%; Score 21.4; DB 20; Length 1306;
Best Local Similarity 95.7%; Pred. No. 0.63;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 cgattgccagcaaatgcgaagt 25
Db 195 cgcttgccagcaaatgcgaagt 217
RESULT 15
AAD05462
ID AAD05462 standard; cDNA; 1312 BP.
XX
XX AAD05462;
XX
DT 17-JUL-2001 (first entry)
XX
DE Human secreted protein-encoding gene 23 cDNA clone HNTRS57, SEQ ID NO:84.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; kidney disorder;
KW cardiovascular disorder; angioinfectious disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulvovaginal;
KW cell culture; chemotaxis; food additive; gene therapy;
KW binding partner identification; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 69..674
FT /*tag= a
FT /product= "Human secreted protein"
FT /note= "CDS does not include start codon"
FT /partial
FT 69..71
FT sig_peptide /*tag= b
FT 72..671
FT mat_peptide /*tag= c
FT /product= "Mature human secreted protein"
XX
XX WO200134623-A1.
XX
XX 17-MAY-2001.
XX
XX 01-NOV-2000; 2000WO-US30037.
XX
XX 05-NOV-1999; 99US-0163577.
XX 30-JUN-2000; 2000US-0215137.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Komatsoulis GA, Moore PA;
XX
XX WPI; 2001-316490/33.
XX P-PSDB; AAE01619.
XX
XX Nucleic acids encoding 29 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
XX Claim 1; Page 469; 535pp; English.
XX
XX AAD05389-AAD05473 represent cDNAs corresponding to 29 human secreted
CC protein genes, and AAE01546-AAE01630 represent the proteins they encode.
CC AAE01631-AAE01660 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,

CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 29 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiotensin disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein-encoding cDNA of the invention.
 XX
 SQ Sequence 1312 BP: 334 A; 288 C; 382 G; 304 T; 4 other;

Query Match 85.6%; Score 21.4; DB 22; Length 1312;
 Best Local Similarity 95.7%; Pred. No. 0.63;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 3 cgattgccagcaaatgcgaagt 25
 |||
 Db 201 cgcttgccagcaaatgcgaagt 223

Search completed: May 24, 2002, 00:23:11
 Job time: 9008 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:12:11 ; Search time 3953.4 Seconds
(without alignments)
132.333 Million cell updates/sec

Title: US-09-696-686-46

Perfect score: 25
Sequence: 1 tgcgattgccagcaaatgcgaagt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	25	100.0	1882	10	BC013549	BC013549 Mus muscu
2	25	100.0	1909	10	AF361644	AF361644 Mus muscu
3	24	96.0	89855	2	AC074220	AC074220 Mus muscu
4	21.4	85.6	510	6	AX079702	AX079702 Sequence
5	21.4	85.6	744	6	AR142811	AR142811 Sequence
6	21.4	85.6	746	6	AX136616	AX136616 Sequence
7	21.4	85.6	747	6	BD006701	BD006701 Novel pol
8	21.4	85.6	1413	6	BC019903	BC019903 Homo sapi
9	21.4	85.6	1419	6	AX136423	AX136423 Sequence
10	21.4	85.6	1512	6	AR142810	AR142810 Sequence
11	20.4	81.6	128185	2	AC087144	AC087144 Mus muscu
12	20.4	81.6	132764	2	AC073842	AC073842 Homo sapi
13	20.4	81.6	167854	2	AC092483	AC092483 Homo sapi
14	20.4	81.6	209958	2	AC087135	AC087135 Mus muscu
15	20.2	80.8	962	9	HS080744	U00744 Homo sapien
16	20.2	80.8	1422	9	BC008898	BC008898 Homo sapi
17	20.2	80.8	1422	9	BC008961	BC008961 Homo sapi
18	20.2	80.8	1720	9	BC008961	BC008961 Homo sapi
19	20.2	80.8	190043	9	BC008961	BC008961 Homo sapi
20	19.2	76.8	113109	9	HS475N16	AL035587 Human DNA
21	19.2	76.8	160609	2	AC096211	AC096211 Rattus no
22	18.8	75.2	1363	9	HUMBINDA	L10403 Homo sapien
23	18.8	75.2	2008	9	AK023954	AK023954 Homo sapi
24	18.8	75.2	2096	6	AX192223	AX192223 Sequence
25	18.8	75.2	3492	10	AF175967	AF175967 Mus muscu
26	18.8	75.2	4349	10	AF175968	AF175968 Mus muscu
27	18.8	75.2	154508	2	AC083974	AC083974 Homo sapi
28	18.8	75.2	160741	9	AC097709	AC097709 Homo sapi
29	18.8	75.2	168741	2	AC013573	AC013573 Homo sapi
30	18.8	75.2	16891	2	AC106523	AC106523 Rattus no
31	18.8	75.2	178608	2	AC027531	AC027531 Homo sapi
32	18.8	75.2	193625	2	AP001194	AP001194 Homo sapi
33	18.8	75.2	280810	2	AC046188	AC046188 Homo sapi
34	18.6	74.4	2029	10	AF184981	AF184981 Mus muscu
35	18.6	74.4	185893	9	AC005191	AC005191 Homo sapi
36	18.2	72.8	3191	9	BC011454	BC011454 Homo sapi
37	18.2	72.8	3542	9	AF175966	AF175966 Homo sapi
38	18.2	72.8	4412	9	AB023206	AB023206 Homo sapi
39	18.2	72.8	8509	2	AC012985	AC012985 Drosophila
40	18.2	72.8	28928	3	CEZC410	268270 Caenorhabdi
41	18.2	72.8	45978	3	AF016449	AF016449 Caenorhab
42	18.2	72.8	74484	2	AC106387	AC106387 Rattus no
43	18.2	72.8	92969	2	AC010844	AC010844 Drosophila
44	18.2	72.8	113635	2	AC108252	AC108252 Rattus no
45	18.2	72.8	117136	2	AC015940	AC015940 Homo sapi

ALIGNMENTS

RESULT 1

BC013549

LOCUS

DEFINITION

BC013549

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC013549 Mus musculus, RIKEN cDNA 1600025D17 gene, clone MGC:19340
IMAGE:4222133, mRNA, complete cds.
1882 bp mRNA linear ROD 07-SEP-2001

BC013549 GI:15488835

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1882)

Strausberg R.

Direct Submission

Submitted (04-SEP-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: cgapps-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

Fri May 24 10:49:39 2002

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdspaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 25 Row: j Column: 9.

FEATURES
 source
 1..1882
 /organism="Mus musculus"
 /db_xref="LocusID:72029"
 /db_xref="taxon:10090"
 /clone="MGC:19340 IMAGE:4222133"
 /tissue="kidney, normal, 5 month old male mouse."
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 91..921
 /codon_start=1
 /product="RIKEN cDNA 1600025D17 gene"
 /protein_id="AAH13549.1"
 /db_xref="GI:15488836"
 /translation="MESSELAAPRCLLPFLLLPAPKLGSPAGAEETDWR
 LPSCVEYKVAVELKSAFETGKTEVIDTGYGLDGGKGVKTKSLRLIEVTEF
 ICKRLDYLHRTGSRNFAKGMSETFETLHNLVHKGKVMVDIPYELWNETSAEVA
 DLKKQCDVLVEFEFIEDWYRNHQQEEDTEFLCANHLKGDTSCLAERWSGKGGDI
 ASLGKSKKKRSGVSGSSGSKKORKEGLGLEDANAEPEGVQKASPLPHSPDDEL
 "

CDS
 432 a 539 c 549 g 362 t
 BASE COUNT
 ORIGIN
 Query Match 100.0%; Score 25; DB 10; Length 1882;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgcgattgccagcaaatgcgaagt 25
 |||||
 Db 218 TCGATTGCCAGCAAAATGCGAAGT 242
 |||||
 RESULT 2
 AF361644 1909 bp mRNA linear ROD 12-MAY-2001
 LOCUS
 DEFINITION
 Mus musculus putative retinoic acid-regulated protein mRNA,
 complete cds.
 ACCESSION
 AF361644
 VERSION
 AF361644.1 GI:14028766
 KEYWORDS
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 1909)
 Glozak, M.A., Li, Y., Reuille, R. and Rogers, M.B.
 Trapping and characterization of novel retinoic acid response
 elements
 JOURNAL
 Unpublished
 2 (bases 1 to 1909)
 Glozak, M.A., Li, Y., Reuille, R. and Rogers, M.B.
 AUTHORS
 Direct Submission
 TITLE
 Submitted (15-MAR-2001) Biology, University of South Florida, 4202
 E. Fowler Avenue, Tampa, FL 33620, USA
 JOURNAL
 Location/Qualifiers
 FEATURES
 source
 1..1909
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /cell_line="P19; derived from matings between C3H/He

females and male animals carrying an X chromosome derived
 from a feral mouse (McBurney & Rogers, Dev. Biol., 89,
 503)"
 /cell_type="embryonal carcinoma"
 /clone_lib="24 hr retinoic acid-treated P19 cDNA (Bouillet
 et al., Dev. Biol., 170, 420)"
 89..919
 CDS
 /codon_start=1
 /product="putative retinoic acid-regulated protein"
 /protein_id="AAK52494.1"
 /db_xref="GI:14028767"
 /translation="MESSELAAPRCLLPFLLLPAPKLGSPAGAEETDWR
 LPSCVEYKVAVELKSAFETGKTEVIDTGYGLDGGKGVKTKSLRLIEVTEF
 ICKRLDYLHRTGSRNFAKGMSETFETLHNLVHKGKVMVDIPYELWNETSAEVA
 DLKKQCDVLVEFEFIEDWYRNHQQEEDTEFLCANHLKGDTSCLAERWSGKGGDI
 ASLGKSKKKRSGVSGSSGSKKORKEGLGLEDANAEPEGVQKASPLPHSPDDEL
 "

BASE COUNT 424 a 536 c 558 g 386 t 5 others
 ORIGIN
 Query Match 100.0%; Score 25; DB 10; Length 1909;
 Best Local Similarity 100.0%; Pred. No. 0.047;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 tgcgattgccagcaaatgcgaagt 25
 |||||
 Db 216 TCGATTGCCAGCAAAATGCGAAGT 240
 |||||

RESULT 3
 AC074220/c 89855 bp DNA linear HTG 25-JAN-2002
 LOCUS
 DEFINITION
 Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
 WORKING DRAFT SEQUENCE, 38 unordered pieces.
 ACCESSION
 AC074220
 VERSION
 AC074220.7 GI:13876390
 KEYWORDS
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 89855)
 Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
 Toshihikesh, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
 Gordon, M., Goltz, J.S. and Kucherlapati, R.
 High Throughput Mouse Sequencing
 Unpublished
 2 (bases 1 to 89855)
 Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
 Toshihikesh, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
 Gordon, M., Goltz, J.S. and Kucherlapati, R.
 Direct Submission
 TITLE
 Submitted (20-JUL-2000) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA
 On Apr 28, 2001 this sequence version replaced gi:11276104.
 COMMENT
 -----Genome Center
 Center: Harvard Partners Genome Center
 Center Code: HPGC
 Web site:
<http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous>
 e.html
 Contact: gnktm@cecod.bwh.harvard.edu
 -----Summary Statistics
 Center project name: ADY
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100%
 *Consensus quality: 79355 at least Q20
 *Consensus quality: 75365 at least Q30
 *Consensus quality: 69022 at least Q40
 Estimated insert size: agarose-FP - N/A
 **Estimated insert size: 89115 - sum-of-contigs

Quality coverage: agarose-FP - N/A
Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	7962: contig of 7962 bp in length
7963	7982: gap of unknown length
7983	13924: contig of 5942 bp in length
13925	13944: gap of unknown length
13945	10990: contig of 5046 bp in length
18991	19010: gap of unknown length
19011	22501: contig of 3491 bp in length
22502	22521: gap of unknown length
22522	27077: contig of 4556 bp in length
27078	27097: gap of unknown length
27098	28819: contig of 1722 bp in length
28820	28839: gap of unknown length
28840	31224: contig of 2385 bp in length
31225	31244: gap of unknown length
31245	33093: contig of 3849 bp in length
35094	35113: gap of unknown length
35114	38611: contig of 3498 bp in length
38612	38631: gap of unknown length
38632	42261: contig of 3630 bp in length
42262	42281: gap of unknown length
42282	44130: contig of 1849 bp in length
44131	44150: gap of unknown length
44151	47053: contig of 2909 bp in length
47060	47079: gap of unknown length
47080	49496: contig of 2417 bp in length
49497	49516: gap of unknown length
49517	52236: contig of 2720 bp in length
52237	52256: gap of unknown length
52257	54270: contig of 2014 bp in length
54271	54290: gap of unknown length
54291	56271: contig of 1981 bp in length
56272	56291: gap of unknown length
56292	58849: contig of 2558 bp in length
58850	58869: gap of unknown length
58870	60631: contig of 1752 bp in length
60622	60641: gap of unknown length
60842	62255: contig of 1614 bp in length
62256	62275: gap of unknown length
62276	63085: contig of 810 bp in length
63086	63105: gap of unknown length
63106	65440: contig of 2335 bp in length
65441	65460: gap of unknown length
65461	67629: contig of 2169 bp in length
67630	67649: gap of unknown length
67650	69178: contig of 1529 bp in length
69179	69198: gap of unknown length
69199	70095: contig of 897 bp in length
70096	70115: gap of unknown length
70116	71821: contig of 1706 bp in length
71822	71841: gap of unknown length
71842	73506: contig of 1665 bp in length
73507	73526: gap of unknown length
73527	74749: contig of 1223 bp in length
74750	74769: gap of unknown length
74770	75963: contig of 1194 bp in length
75964	75983: gap of unknown length
75984	77905: contig of 1922 bp in length
77906	77925: gap of unknown length
77926	79236: contig of 1311 bp in length
79237	79256: gap of unknown length
79257	81332: contig of 2076 bp in length
81333	81352: gap of unknown length

FEATURES

source

Location/Qualifiers

```

Location/Quailers
1. .89855
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-76116"

```

misc feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature		73527..74749	
		/note="assembly_name:Contigl19"	
misc_feature		74770..75963	
		/note="assembly_name:Contigl18"	
misc_feature		75984..77905	
		/note="assembly_name:Contigl17"	
misc_feature		77926..79236	
		/note="assembly_name:Contigl16"	
misc_feature		79257..81332	
		/note="assembly_name:Contigl15"	
misc_feature		81353..83169	
		/note="assembly_name:Contigl14"	
misc_feature		83190..84096	
Query Match		96.0%; Score 24; DB 2; Length 89855;	
Best Local Similarity		100.0%; Pred. No. 0.25;	
Matches 24; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY		1 tgcgattgccagcaaatgcgaag 24	
Db		5896 TCGATTGCCAGCAAAATGCGAAG 5873	
RESULT		4	
LOCUS		AX079702 510 bp DNA linear PAT 22-FEB-2001	
DEFINITION		Sequence 446 from Patent WO0107611.	
ACCESSION		AX079702	
VERSION		AX079702.1 GI:13159271	
KEYWORDS		human.	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1 (bases 1 to 510)	
AUTHORS		Baker,K.P., Goddard,A. and Wood,W.I.	
TITLE		Human polypeptides and methods for the use thereof	
JOURNAL		Patent: WO 0107611-A 446 01-FEB-2001;	
FEATURES		Genentech, Inc. (US)	
source		Location/Qualifiers	
		1..510	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
BASE COUNT		122 a 114 c 168 g 104 t	
ORIGIN		2 others	
Query Match		85.6%; Score 21.4; DB 6; Length 510;	
Best Local Similarity		95.7%; Pred. No. 3.2;	
Matches 22; Conservative		0; Mismatches 1; Indels 0; Gaps 0;	
QY		3 cgattgccagcaaatgcgaagt 25	
Db		206 CGTTGCCAGCAAAATGCGAAGT 228	
RESULT		5	
LOCUS		AR142811 744 bp DNA linear PAT 08-AUG-2001	
DEFINITION		Sequence 3 from patent US 6204013.	
ACCESSION		AR142811	
VERSION		AR142811.1 GI:15104097	
KEYWORDS		Unknown.	
SOURCE		Unknown.	
ORGANISM		Unclassified.	
REFERENCE		1 (bases 1 to 744)	
AUTHORS		Khodadoust,M.M.	
TITLE		MSP-5 nucleic acid molecules and uses therefor	
JOURNAL		Patent: US 6204013-A 3 20-MAR-2001;	
FEATURES		Location/Qualifiers	
source		1..744	
		/organism="unknown"	
BASE COUNT		122 a 114 c 168 g 104 t	
ORIGIN		2 others	
Query Match		85.6%; Score 21.4; DB 6; Length 510;	
Best Local Similarity		95.7%; Pred. No. 3.2;	
Matches 22; Conservative		0; Mismatches 1; Indels 0; Gaps 0;	
QY		3 cgattgccagcaaatgcgaagt 25	
Db		206 CGTTGCCAGCAAAATGCGAAGT 228	
RESULT		7	
LOCUS		BD006701 747 bp DNA linear PAT 31-JAN-2002	
DEFINITION		Novel polypeptide.	
ACCESSION		BD006701	
VERSION		BD006701.1 GI:18635072	
KEYWORDS		JP 2001029090-A/4.	
SOURCE		Homo sapiens	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		Ito,Y., Mogi,S., Tanaka,H., Okubo,S. and Ogi,K.	
TITLE		Novel polypeptide	
JOURNAL		Patent: JP 2001029090-A 4 06-FEB-2001;	
COMMENT		TAKEDA CHEMICAL INDUSTRIES LTD	
		OS Homo sapiens (human)	
		PN JP 2001029090-A/4	
		PD 06-FEB-2001	
		PF 19-MAY-2000 JP 2000147530	
		PR	
		PI YASUAKI ITO, SHINTACHI MOGI, HIDEYUKI TANAKA, SHOICHI OKUBO, PI	
		KAZUHIRO OGI	
		PC C12N15/09, A61K38/00, A61K39/395, A61K45/00, A61K48/00,	
		PC A61P1/00,	
		PC A61P5/00, A61P9/00, A61P11/00, A61P25/00, A61P31/00, PC	


```

A61P35/00,
PC A61P37/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,G01N33/50,
PC G01N33/56//
PC C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/08,(C12P21/02, PC
C12R1:91),
PC (C12N5/10,C12R1:91),(C12P21/08,C12R1:91),C12N15/00,A61K37/02,
PC C12N5/00,
CC (C12N5/00,C12R1:91)
FH Key Location/Qualifiers
FT source 1..747
FT /organism="Homo sapiens (human)".
FEATURES
    source
        Location/Qualifiers
            1..747
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
BASE COUNT 209 a 143 c 255 g 140 t
ORIGIN
Query Match 85.6%; Score 21.4; DB 6; Length 747;
Best Local Similarity 95.7%; Pred. No. 3.4;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 cgattgccagcaaatgcgaagt 25
|| |||||
Db 97 CGCTTGCCCAAGTGCAGT 119

RESULT 8
BC019903 1413 bp mRNA linear PRI 22-JAN-2002
LOCUS Homo sapiens, clone IMAGE:4991480, mRNA, partial cds.
DEFINITION BC019903
ACCESSION BC019903
VERSION BC019903.1 GI:18043798
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1413)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Guanatne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 42 Row: j Column: 23.
    Location/Qualifiers
        1..1413
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4991480"
            /tissue_type="Placenta, choriochorinoma"
            /clone_id="NIH_MGC_10"
            /lab_host="DH10B"

```

```

/note="Vector: PCMV-SPORT6"
<1..774
/codon_start=1
/product="Unknown (protein for IMAGE:4991480)"
/protein_id="AAH19903.1"
/db_xref="GI:18043798"
/translation="OAVTCRRRRVMPVRLGILLFLFLAVHEAWAGMKKEDDDDTERLPKSCVEYCKLL
LPSKCEVCKLLSTELQALSTGRSREVLEQGLDVGKVRKHVPYSVSETRLEALE
NLCEIRLDYSVHAERKSLRYAKGOSQTMATLKGVLQKGVKVDLGIPLWDEPSVEY
TYLKKQCEITMLEFEDIVGDWYFHQEQPLQNFCEGHVLPAAETACIQETWTGKEIT
DGEKTEGEEERQEEEEEGGDKMTKTSHPKLDREDL"
BASE COUNT 378 a 301 c 413 g 321 t
ORIGIN

Query Match 85.6%; Score 21.4; DB 9; Length 1413;
Best Local Similarity 95.7%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 cgattgccagcaaatgcgaagt 25
|| |||||
Db 130 CGCTTGCCCAAGTGCAGT 152

RESULT 9
AX136423 1419 bp DNA linear PAT 30-MAY-2001
LOCUS AX136423
DEFINITION Sequence 345 from Patent EP1067182.
ACCESSION AX136423
VERSION AX136423.1 GI:14272827
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1419)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
Hayashi,K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 345 10-JAN-2001;
Helix Research Institute (JPI)
FEATURES
    source
        Location/Qualifiers
            1..1419
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAC39850.1"
                /db_xref="GI:14272828"
                /translation="MGVRLGILLFLFLAVHEAWAGMKKEDDDDTERLPKSCVEYCKLL
                STELOAELSTGRSREVLEQGLDVGKVRKHVPYSVSETRLEALENLCEIRLDYSV
                HAERKSLRYAKGOSQTMATLKGVLQKGVKVDLGIPLWDEPSVEYTKQCEITL
                EFEDIVGDWYFHQEQPLQNFCEGHVLPAAETACIQETWTGKEITDGEKTEGEE
                EEEEEEGGDKMTKTSHPKLDREDL"
BASE COUNT 363 a 307 c 422 g 327 t
ORIGIN

Query Match 85.6%; Score 21.4; DB 6; Length 1419;
Best Local Similarity 95.7%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 cgattgccagcaaatgcgaagt 25
|| |||||
Db 154 CGCTTGCCCAAGTGCAGT 176

RESULT 10
AR142810 1512 bp DNA linear PAT 08-AUG-2001
LOCUS AR142810
DEFINITION Sequence 1 from patent US 6204013.
ACCESSION AR142810

```

* 1120	2601: contig of 1482 bp in length
* 2602	2701: gap of unknown length
* 2702	4040: contig of 1339 bp in length
* 4041	4140: gap of unknown length
* 4141	5695: contig of 1555 bp in length
* 5696	5795: gap of unknown length
* 5796	8073: contig of 2278 bp in length
* 8074	8173: gap of unknown length
* 8174	11247: contig of 3074 bp in length
* 11248	11347: gap of unknown length
* 11348	13077: contig of 1730 bp in length
* 13078	13177: gap of unknown length
* 13178	17301: contig of 4124 bp in length
* 17302	17401: gap of unknown length
* 17402	21914: contig of 4513 bp in length
* 21915	22014: gap of unknown length
* 22015	30777: contig of 8763 bp in length
* 30778	30877: gap of unknown length
* 30878	43650: contig of 12773 bp in length
* 43651	43750: gap of unknown length
* 43751	56081: contig of 12331 bp in length
* 56082	56181: gap of unknown length
* 56182	80044: contig of 23863 bp in length
* 80045	80144: gap of unknown length
* 80145	128185: contig of 48041 bp in length
FEATURES	
Source	
1. .128185	
/organism="Mus musculus"	
/db_xref="taxon:10090"	
/clone_lib="RP23-257A7"	
/clone_lib="RPC1 mouse BAC library 23"	
BASE COUNT	33237 a 30255 c 29682 g 33675 t 1336 others
ORIGIN	
Query Match 81.6%; Score 20.4; DB 2; Length 128185;	
Best Local Similarity 95.5%; Pred. No. 21;	
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	3 cgattgccagcaaatgcgaag 24
Db	116457 CGTTGCCAGCAATGCGAAG 116478
RESULT 12	
AC073842/c	132764 bp DNA linear PRI 09-JAN-2002
LOCUS	AC073842
DEFINITION	Homo sapiens BAC clone RP11-506M12 from 7, complete sequence.
ACCESSION	AC073842
VERSION	AC073842.5 GI:18095607
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 132764)
AUTHORS	Sulston,J.E. and Waterston,R.
TITLE	Toward a complete human genome sequence
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE	99063792
REFERENCE	2 (bases 1 to 132764)
AUTHORS	Isak,A., Kozlowicz,A. and Boyer,E.
TITLE	The sequence of Homo sapiens BAC clone RP11-506M12
JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 132764)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (29-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
MO 63108, USA	
REFERENCE	4 (bases 1 to 132764)
AUTHORS	Waterston,R.
TITLE	Direct Submission

JOURNAL

Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:114647315.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_NH0506M12

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgrl.nih.gov/DIR/GR7>, send mailto:egreen@nhgrl.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-161A2, 2000 bp overlap; the clone sequenced to the right is RP13-741A20. Actual end is at base position 132764 of RP11-506M12.

Data from AC093619 was used to finish this clone, AC073842.

Polymorphisms have been identified between AC005071 and AC073842.

FEATURES

source

Location/Qualifiers

1. .132764
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone_lib="RPCI-11"
/clone="rp11-506M12"

misc_feature

1. .397
/note="similar to EST AA009576 (NID:gl1470735) ze82d10.r1"

misc_feature

1. .397

misc_feature

1. .67

misc_feature

1. .67

misc_feature

1. .67

misc_feature

2. .397

repeat_region

462. .513

/rpt_family="L2"
572. .673
/note="similar to EST BE729033 (NID:gl10143025)"
572. .673
/note="similar to EST BF026637 (NID:gl10734349)"
727. .1020
/note="similar to EST AA077273 (NID:gl1836747)"
881. .1191
/rpt_family="Alu"
1165. .1206
/rpt_family="(TAAA)n"
1214. .1520
/rpt_family="Alu"
1495. .1522
/rpt_family="(GAAA)n"
1603. .1723
/rpt_family="L2"
2050. .2494
/note="similar to EST AW953106 (NID:gl142789)"
2090. .2242
/note="similar to EST T08560 (NID:gl389588)"
2104. .2240
/note="similar to EST BE666576 (NID:gl10027167)"
2104. .2240
/note="similar to EST BF026637 (NID:gl10734349)"
2104. .2240
/note="similar to EST BE729033 (NID:gl10143025)"
2107. .2493
/note="similar to EST R85198 (NID:gl943604) y041a11.r1"
2140. .2240
/note="similar to EST AW258515 (NID:gl6631496) uq36b09.y1"
2298. .3060
/note="similar to EST AV723000 (NID:gl0826061)"
2347. .2832
/note="similar to EST H18846 (NID:gl985086) ym45e12.r1"
2347. .2522
/note="similar to EST BE534256 (NID:gl9762901)"
2354. .2462
/note="similar to EST BE666576 (NID:gl10027167)"
2354. .2522
/note="similar to EST BE729033 (NID:gl10143025)"
2354. .2522
/note="similar to EST AW258515 (NID:gl6631496) uq36b09.y1"
2354. .2522
/note="similar to EST BF026637 (NID:gl10734349)"
2359. .2522
/note="similar to EST BE727269 (NID:gl10141362)"
2407. .2522
/note="similar to EST BE664898 (NID:gl10024867)"
2478. .2837
/note="similar to EST BF510418 (NID:gl11593716)"
2533. .3040
/note="similar to EST H17220 (NID:gl983460) ym42c05.r1"
2535. .3115
/note="similar to EST T75205 (NID:gl691967) yc86e12.r1"
2653. .2725
/note="similar to EST BE729033 (NID:gl10143025)"
2653. .2745
/note="similar to EST AW258515 (NID:gl6631496) uq36b09.y1"
2653. .2745
/note="similar to EST BE664898 (NID:gl10024867)"
2653. .2725
/note="similar to EST BE727269 (NID:gl10141362)"
2653. .2725
/note="similar to EST BF026637 (NID:gl10734349)"
2653. .2745
/note="similar to EST BE534256 (NID:gl9762901)"
2742. .2987
/note="similar to EST Z42144 (NID:gl564385)"
2800. .3015
/note="similar to EST BI037302 (NID:gl14443928)"
2828. .2908
/note="similar to EST BE664898 (NID:gl10024867)"

```

misc_feature 2828. 2908
/note="similar to EST AW258515 (NID:g6631496) uq36b09.y1"
misc_feature 2828. 2908
/note="similar to EST BE534256 (NID:g9762901)"
misc_feature 2863. 3309
/note="similar to EST R54265 (NID:g816167) yg74h03.rl"
misc_feature 2893. 3290
/note="similar to EST H14061 (NID:g878909) ym63a03.s2"
misc_feature 2905. 3554
/note="similar to EST BF921437 (NID:g12317325)"
misc_feature 3343. 3430
/note="similar to EST BE534256 (NID:g9762901)"
misc_feature 3343. 3430
/note="similar to EST BE664898 (NID:g10024867)"
misc_feature 3343. 3428
/note="similar to EST AW258515 (NID:g6631496) uq36b09.y1"
misc_feature 3434. 3950
/note="similar to EST AA169431 (NID:g1748398) zo84g01.rl"
misc_feature 3487. 4034
/note="similar to EST BF939888 (NID:g12357208)"
misc_feature 3506. 3669
/note="similar to EST BE073223 (NID:g8420006)"
misc_feature 3534. 3701
/note="similar to EST BE664898 (NID:g10024867)"
misc_feature 3549. 4041
/note="similar to EST H17115 (NID:g883355) ym42c05.sl"
misc_feature 3681. 4041

```

```

Query Match 81.68; Score 20.4; DB 9; Length 132764;
Best Local Similarity 95.5%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 cgattgccagcaaatggaag 24
|||
Db 84853 GCCTGCCAGCAATGCGAAG 84832

```

```

RESULT 13
AC092483/3 167854 bp DNA linear HTG 21-AUG-2001
LOCUS Homo sapiens chromosome UNK clone CTD-2188L12, *** SEQUENCING IN
DEFINITION PROGRESS ***
AC092483 AC092483.2 GI:15217361
VERSION HTG; HTGS_PHASE1.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167854)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 167854)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 21, 2001 this sequence version replaced gi:14647307.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.MS2188L12
----- Summary Statistics -----
Sequencing vector: M13; 92%
Sequencing vector: plasmid; 8%
Chemistry: Dye-terminator ET; 3% of reads
Chemistry: Dye-terminator Big Dye; 97% of reads
Assembly program: Phrap; version 0.990319

```

```

Consensus quality: 115486 bases at least Q40
Consensus quality: 134295 bases at least Q30
Consensus quality: 146715 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 163554; sum-of-contigs
Quality coverage: 2.74 in Q20 bases; agarose-fp
Quality coverage: 2.79 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 1021: contig of 1021 bp in length
* 1022 1121: gap of unknown length
* 1122 2724: contig of 1603 bp in length
* 2725 2824: gap of unknown length
* 2825 3877: contig of 1053 bp in length
* 3878 3977: gap of unknown length
* 3978 5593: contig of 1616 bp in length
* 5594 5694: gap of unknown length
* 5695 8414: contig of 2721 bp in length
* 8415 8514: gap of unknown length
* 8515 9718: contig of 1204 bp in length
* 9719 9818: gap of unknown length
* 9819 12515: contig of 2697 bp in length
* 12516 12616: gap of unknown length
* 12617 14391: contig of 1775 bp in length
* 14392 14490: gap of unknown length
* 14491 17360: contig of 2870 bp in length
* 17361 17461: gap of unknown length
* 17462 19531: contig of 2071 bp in length
* 19532 19632: gap of unknown length
* 19633 22131: contig of 2500 bp in length
* 22132 22232: gap of unknown length
* 22233 23694: contig of 1462 bp in length
* 23695 23794: gap of unknown length
* 23795 26183: contig of 2389 bp in length
* 26184 26282: gap of unknown length
* 26283 28149: contig of 1867 bp in length
* 28150 28249: gap of unknown length
* 28250 30016: contig of 1767 bp in length
* 30017 30117: gap of unknown length
* 30118 31897: contig of 1781 bp in length
* 31898 31997: gap of unknown length
* 31998 34152: contig of 2155 bp in length
* 34153 34252: gap of unknown length
* 34253 36342: contig of 2090 bp in length
* 36343 36442: gap of unknown length
* 36443 40089: contig of 3647 bp in length
* 40090 40189: gap of unknown length
* 40190 42653: contig of 2463 bp in length
* 42654 42752: gap of unknown length
* 42753 45366: contig of 2614 bp in length
* 45367 45466: gap of unknown length
* 45467 48030: contig of 2564 bp in length
* 48031 48130: gap of unknown length
* 48131 51385: contig of 3255 bp in length
* 51386 51486: gap of unknown length
* 51487 55007: contig of 3522 bp in length
* 55008 55107: gap of unknown length
* 55108 59428: contig of 4321 bp in length
* 59429 59528: gap of unknown length
* 59529 63903: contig of 4375 bp in length
* 63904 64003: gap of unknown length
* 64004 68557: contig of 4554 bp in length
* 68558 68657: gap of unknown length
* 68659 74048: contig of 5391 bp in length
* 74049 74148: gap of unknown length
* 74149 78748: contig of 4600 bp in length

```

```

* 78749      78848: gap of unknown length.
* 78849      83676: contig of 4828 bp in length.
* 83676      83776: gap of unknown length.
* 83777      87313: contig of 3537 bp in length.
* 87314      87413: gap of unknown length.
* 87414      91962: contig of 4549 bp in length.
* 91963      92062: gap of unknown length.
* 92063      97831: contig of 5769 bp in length.
* 97832      97931: gap of unknown length.
* 97932      102367: contig of 4336 bp in length.
* 102368     102367: gap of unknown length.
* 102368     102368: contig of 5134 bp in length.
* 102368     107501: gap of unknown length.
* 107502     107502: contig of 4984 bp in length.
* 107502     112585: gap of unknown length.
* 112586     112586: gap of unknown length.
* 112586     117771: contig of 5086 bp in length.
* 117772     117772: gap of unknown length.
* 117872     123681: contig of 5810 bp in length.
* 123682     123781: gap of unknown length.
* 123782     129137: contig of 5356 bp in length.
* 129138     129237: gap of unknown length.
* 129238     135973: contig of 6736 bp in length.
* 135974     136073: gap of unknown length.
* 136074     144530: contig of 8557 bp in length.
* 144631     144730: gap of unknown length.
* 144731     152830: contig of 8100 bp in length.
* 152831     152930: gap of unknown length.
* 152931     161016: contig of 8086 bp in length.
* 161017     161116: gap of unknown length.
* 161117     167854: contig of 6738 bp in length.

FEATURES             Location/Qualifiers
    source            1..167854
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="UNK"
                    /clone="CTD-2188L12"
misc_feature          1..1021
                    /note="assembly_name:Contig20"
misc_feature          1122..2724
                    /note="assembly_name:Contig21"
misc_feature          2825..3877
                    /note="assembly_name:Contig22"
misc_feature          3978..5593
                    /note="assembly_name:Contig23"
misc_feature          5694..8414
                    /note="assembly_name:Contig24"
misc_feature          8515..9718
                    /note="assembly_name:Contig25"
misc_feature          9819..12515
                    /note="assembly_name:Contig26"
misc_feature          12616..14390
                    /note="assembly_name:Contig27"
misc_feature          14491..17360
                    /note="assembly_name:Contig28"
misc_feature          17461..19531
                    /note="assembly_name:Contig29"
misc_feature          19632..22131
                    /note="assembly_name:Contig30"
misc_feature          22232..23693
                    /note="assembly_name:Contig31"
misc_feature          23794..26182
                    /note="assembly_name:Contig32"
misc_feature          26283..28149
                    /note="assembly_name:Contig33"
misc_feature          28250..30016
                    /note="assembly_name:Contig34"
misc_feature          30117..31897
                    /note="assembly_name:Contig35"
misc_feature          31998..34152
                    /note="assembly_name:Contig36"
misc_feature          34253..36342
                    /note="assembly_name:Contig37"
misc_feature          36443..40089

```

```

/note="assembly_name:Contig38"
40190..42652
/note="assembly_name:Contig39"
42753..45366
/note="assembly_name:Contig40"
45467..48030
/note="assembly_name:Contig41"
48131..51385
/note="assembly_name:Contig42"
51486..55007
/note="assembly_name:Contig43"
55108..59428
/note="assembly_name:Contig44"
59529..63903
/note="assembly_name:Contig45"
64004..68557
/note="assembly_name:Contig46"
68658..74048
/note="assembly_name:Contig47"
74149..78748

Query Match           81.6%; Score 20.4; DB 2; Length 167854;
Best Local Similarity 95.5%; Pred.No. 21;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY   3 cgattgccagcaaatgcgaag 24
      |||iiiiiiiii|||
Db   41488 CGCTTCCCGCAGCAATGCGAAG 41467

RESULT 14
AC087135      AC087135      209958 bp      DNA      linear      HTG 09-DEC-2000
LOCUS         Mus musculus clone RP23-147H4, WORKING DRAFT SEQUENCE, 31 unordered
DEFINITION    pieces.
ACCESSION     AC087135
VERSION       AC087135.1 GI:11610861
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 209958)
AUTHORS       DOE Joint Genome Institute.
TITLE         Sequencing of Mouse
JOURNAL        Unpublished
REFERENCE     2 (bases 1 to 209958)
AUTHORS       DOE Joint Genome Institute.
TITLE         Direct Submission
JOURNAL        Submitted (09-DEC-2000) Production Sequencing Facility, DOE Joint
               Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1793492
Center clone name: RPCI-23_147H4
-----
Summary Statistics
Consensus quality: 180916 bases at least Q40
Consensus quality: 198060 bases at least Q30
Consensus quality: 201064 bases at least Q20
Estimated insert size: 204000; agarose-fp estimation
Quality coverage: 8.82 in Q20 bases; agarose-fp estimation
Quality coverage: 8.69 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1017: contig of 1017 bp in length
 * 1018: contig of unknown length
 * 1117: gap of 1366 bp in length
 * 2483: contig of 1366 bp in length
 * 2484: gap of unknown length
 * 2584: contig of 1323 bp in length
 * 3907: gap of unknown length
 * 4006: contig of 1203 bp in length
 * 5209: gap of unknown length
 * 5210: gap of unknown length
 * 5310: contig of 1195 bp in length
 * 6504: gap of unknown length
 * 6605: contig of 1041 bp in length
 * 7645: gap of unknown length
 * 7746: contig of 1430 bp in length
 * 9175: gap of unknown length
 * 9176: gap of unknown length
 * 9276: contig of 1112 bp in length
 * 10387: contig of 1112 bp in length
 * 10487: gap of unknown length
 * 10488: contig of 1199 bp in length
 * 11686: contig of 1199 bp in length
 * 11687: gap of unknown length
 * 11787: contig of 1652 bp in length
 * 13439: gap of unknown length
 * 13539: contig of 1849 bp in length
 * 15388: gap of unknown length
 * 15488: contig of 1667 bp in length
 * 17154: gap of unknown length
 * 17254: contig of 1444 bp in length
 * 17255: gap of unknown length
 * 18698: contig of 1444 bp in length
 * 18699: gap of unknown length
 * 20138: contig of 1340 bp in length
 * 20238: gap of unknown length
 * 20239: contig of 1467 bp in length
 * 21805: gap of unknown length
 * 21806: contig of 1513 bp in length
 * 23319: gap of unknown length
 * 23419: contig of 1984 bp in length
 * 25403: gap of unknown length
 * 25503: contig of 2349 bp in length
 * 27851: gap of unknown length
 * 27852: contig of 3005 bp in length
 * 30956: gap of unknown length
 * 30957: contig of 5566 bp in length
 * 31057: gap of unknown length
 * 36623: contig of 7489 bp in length
 * 36723: gap of unknown length
 * 44212: contig of 10212 bp in length
 * 44312: gap of unknown length
 * 44313: contig of 6544 bp in length
 * 54524: gap of unknown length
 * 54624: contig of 7401 bp in length
 * 61168: gap of unknown length
 * 61288: contig of 7401 bp in length
 * 68668: gap of unknown length
 * 68669: contig of 9978 bp in length
 * 68769: gap of unknown length
 * 78747: contig of 17929 bp in length
 * 78847: gap of unknown length
 * 96775: gap of unknown length
 * 96776: contig of 8907 bp in length
 * 96876: gap of unknown length
 * 105783: contig of 21518 bp in length
 * 105883: gap of unknown length
 * 127401: contig of 11039 bp in length
 * 127501: gap of unknown length
 * 138540: contig of 11039 bp in length
 * 138540: gap of unknown length
 * 138640: contig of 28744 bp in length
 * 167384: gap of unknown length
 * 167484: contig of 42475 bp in length.

FEATURES
 source

1. .209958
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-147H4"
 /clone_lib="RPC1 mouse BAC library 23"
 58060 a 46875 c 46167 g 55804 t 3052 others

Query Match 81.6%; Score 20.4; DB 2; Length 209958;
 Best Local Similarity 95.5%; Pred. No. 22;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgcccagcaaatgcgaag 24
 |||||
 Db 206510 GCTTGGCCCAAGCAATGCGAAG 206531

RESULT 15
 HUS0744 HUS0744 962 bp mRNA linear PRI 18-DEC-1997
 LOCUS Homo sapiens CTG4a mRNA, complete cds.
 DEFINITION U80744
 ACCESSION U80744
 VERSION U80744.1 GI:2565062
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 962)
 AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
 Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
 TITLE cDNAs with long CAG trinucleotide repeats from human brain
 JOURNAL Hum. Genet. 100 (1), 114-122 (1997)
 MEDLINE 97369492

REFERENCE 2 (bases 1 to 962)
 AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
 Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
 TITLE Direct Submission
 JOURNAL Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
 Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA

FEATURES
 source
 1. .962
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /tissue_type="cerebral cortex"

gene
 1. .962
 /gene="CTG4a"
 CDS 388..819
 /gene="CTG4a"
 /note="polyleucine rich"
 /codon_start=1
 /product="CTG4a"
 /protein_id="AAB91442.1"
 /db_xref="GI:2565063"
 /translation="MDSMPAPSRCLLLPLLLLLPLLLPLLLPAPLGPSPQAGAEENDWVR
 LPSCEVKYVAVELKSAFETGKTKEVIGTGYLDQKASGVYTKSDRLRLIETET
 ICKRLIIACTRRGPAIDLPACORPLRHYTTWTGSRW"

repeat_region
 439..462
 /rpt_type=tandem
 /rpt_unit=CTG
 BASE COUNT 200 a 252 c 315 g 194 t 1 others
 ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 962;
 Best Local Similarity 88.0%; Pred. No. 15;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgcgattgcccagcaaatgcgaag 25
 |||||
 Db 515 TTCGCTGCCCAAGCAATGCGAAGT 539

Search completed: May 24, 2002, 00:12:35
 Job time: 11261 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 23, 2002, 23:05:31 ; Search time 4203.75 Seconds
(without alignments)
5904.466 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgagggaggagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1528	83.1	1553	11	AK005532
2	1345.6	73.2	1586	11	AK010702 Mus muscu
3	761.2	41.4	766	10	BG974240 602843972
4	743.8	40.4	793	10	BI104487 602890091
5	722.2	39.3	985	10	BF785328 602111557
6	713	38.8	737	10	BG873094 602794679
7	711.4	38.7	836	10	BI078951 602873203
8	677.8	36.9	942	10	BF385997 602046176
9	646.8	35.2	698	10	BI110829 602895470
10	641.8	34.9	779	10	BI688404 60316338
11	638.2	34.7	645	9	BI498921 60490921
12	637.8	34.7	641	9	BB637089 60637089
13	632.8	34.4	644	9	BB190512 60190512
14	628.4	34.2	656	10	BI082275 602877456
15	626.4	34.1	651	9	BB385488 60385488
16	626	34.0	626	10	BM232538 60324538
17	626	34.0	627	9	BB531047 60531047

18	620.8	33.8	776	10	BG864846
19	619.6	33.7	925	9	AL520942
20	618.4	33.6	1051	10	BM471899
21	613.6	33.4	632	9	BB558801
22	611.8	33.3	656	9	BB504341
23	608	33.1	643	9	BB661603
24	607.4	33.0	625	9	BB633565
25	606.8	33.0	790	10	BF301937
26	606.4	33.0	741	10	BI081472
27	605.6	32.9	829	10	BI833903
28	603.2	32.8	608	9	BB605812
29	599.6	32.6	815	9	AL559536
30	594.2	32.3	665	9	BB218998
31	591.4	32.2	838	10	BF534535
32	586	31.9	916	10	BE790646
33	583	31.7	812	10	BI918555
34	574.6	31.2	808	10	BG686355
35	573.8	31.2	893	10	BG685190
36	571.4	31.1	573	9	AA260027
37	568.2	30.9	713	10	BM042021
38	566.2	30.8	1020	10	BF975546
39	553	30.1	553	10	BM223931
40	547	29.7	666	10	BG681796
41	545.6	29.7	733	10	BI764820
42	541	29.4	623	9	BB661026
43	540.6	29.4	669	10	BG866446
44	537.4	29.2	1003	10	BE250104
45	534.8	29.1	649	9	BB652011

ALIGNMENTS

RESULT 1
AK005532 1553 bp mRNA linear HTC 19-JAN-2002
LOCUS Mus musculus adult female placenta cDNA, RIKEN full-length enriched library, clone:1600025D17:homolog to CTG4A, full insert sequence.
DEFINITION AK005532
ACCESSION AK005532.1 GI:12838152
VERSION HTC; CAP trapper.
KEYWORDS Mus musculus (strain:C57BL/6J) adult female placenta cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
SOURCE clone:1600025D17.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kusunagi,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,S., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913

QY 778 aagcagagaagaactgggggctgggggagagatgccaaacgagagagagagaggt 837
 Db 121 AAGCAGAGAAAGAACTGGGGGGCTGGGGGAGGATGCCAAGCCGAGAGAGAGGCT 180
 QY 838 gtgcagaaggatcgccctccacacacagcccccctgatgagctgtgagccagcttagt 897
 Db 181 GTGCAGAAGCATGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGT 240
 QY 898 gtccctgtaatacgaacccctgactcagagcttgaggacacgcagcgcagcgca 957
 Db 241 GTCCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACACGCACGCGCAGCGCA 300
 QY 958 gtccagcaagagacagctgtctccagcatcaggtctcctcctctggtgtgtccctttc 1017
 Db 301 GCTCCAGCAAGACAGCTGTCTCCAGCATCAGGTCTCTCTCCCTTGGCTGTGCCCTTTC 360
 QY 1018 cttcccttgaaacacagcaaggggtggaagatctggggtgctgggagcgcaccccaa 1077
 Db 361 CTTCCCTTGAACACAGCAAGAGGTGGAAGGATCTGGGTGCTGGGAGACGCCACCCAA 420
 QY 1078 aggaag 1137
 Db 421 AGGAAG 480
 QY 1138 cggggtccaccagcatcccccaggtgagatccaggtcctcctgacatgaagctgaagac 1197
 Db 481 CGGGGTCCACCCAGGATCCCCAGGCTGAGATCCAGGCTTCCAGCATGAAGCTGAAGAGC 540
 QY 1198 atgaggacataaagatgctaccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1257
 Db 541 ATGAGGCACATAAAGATGCTACCCAGGCTGAGATCCAGGCTTCCAGCATGAAGCTGAAGAGC 600
 QY 1258 cagcagcagagcctgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1317
 Db 601 CAGCAGCCAGGCGCTCCCTTCTTCCACCAAGCATCTCTCTCTCTCTCTCTCTCTCT 660
 QY 1318 tggtaaatccagaaactccaggacaaactcgggtgtgtggcacaaggggctggagccag 1377
 Db 661 TGGTAATTCAGAACTTCAGGACAACTCGGGGTGGCACAAGGGGCTGGAGCCCGCAG 720
 QY 1378 agcagagccagccag 1423
 Db 721 AGCCAGAGCCAGCCAGAGACTGCAGAGAGGGGACCTGACTTAACCC 766

RESULT 4
 B1104487
 LOCUS 602890091F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5035030 5',
 DEFINITION mRNA sequence.
 ACCESSION B1104487
 VERSION B1104487.1 GI:14555380
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 793)
 NTM-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 Cloning Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
 http://image.llnl.gov
 Plate: L1AM11097 row: f column: 23
 High quality sequence stop: 784.

FEATURES
 source Location/Qualifiers
 1..793
 /organism="Mus musculus"
 /strain="CZECH II (feral)"
 /db_xref="taxon:10090"
 /clone="IMAGE:5035030"
 /clone_lib="NCI_CGAP_Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 /stem_cell_origin=""
 /lab_host="DH10B"
 /note="Organ: Lung; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"
 BASE COUNT 187 a 197 c 262 g 147 t
 ORIGIN
 Query Match 40.4%; Score 743.8; DB 10; Length 793;
 Best Local Similarity 98.7%; Pred. No. 6e-128;
 Matches 781; Conservative 0; Mismatches 7; Indels 3; Gaps 3;
 QY 9 ggaggaagcgccgcgggtccgctctgctgggtcccgctgagccatgagatgcatgctc 68
 Db 6 CGAGGAAGCGCGCGGGTCCGCTCTGCTCCGGGTCCCGCTGGGCCATGGAGTCCATGTC 65
 QY 69 tgaagtcgc 128
 Db 66 TGAGCTCGCGCGCGCGCTCTTATTTCTTTGCTGTGCTGCTTCCGCTGCTGCTCCT 125
 QY 129 tccgtcccccgaagctagcccgagtcgcccgggctgagggagacgcgactgggtgcgatt 188
 Db 126 TCTTCCCCCGAAGCTAGGCCCGAGTACCCCGCGGCTGAGGACCGCATGGGTGCGGATT 185
 QY 189 gccacgaatcggaagtgtgcaagatgctgctgctgagctgaagctggcttttgagga 248
 Db 186 GCCCAGCAATCGGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTGGC-TTTGAGGA 244
 QY 249 aacgggaagaccgaaggaagtgtgacacccgctatggcaccctggagcggaaggctc 308
 Db 245 AACAGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTCTGGACGGAGGGCTC 304
 QY 309 tggagtcgaagtcacacaaagtcgacttacgttaattgaagtcactgagaccatttgcac 368
 Db 305 TGGAGTCAAGTACACCAAGTCGGACTTACGGTTAATTGAAGTCACTGAGACCATTTGCAA 364
 QY 369 gagctcttgactacagcctgcacaaagagagagagagagagagagagagagagagag 428
 Db 365 GAGGCTTCTGACTACAGCTTGCACAGGAGAGGACTGGCAGCAACCGGTTTGCCAAAGG 424
 QY 429 tatgtcggagaccccttgagagcgtgcacaaactagtcacacaaaggggtcaggggtggtgat 488
 Db 425 TATGTCGGAGACCTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAAGGTTGTTGAT 484
 QY 489 ggaatccccctatgagctgtggaacgagacctcagcagaggtggctgacctcaagaagca 548
 Db 485 GGATATCCCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAACCA 544
 QY 549 gtgtgaagtcgtgtgtggaagagtttgaagaggtgattgagagctggtacaggaaccacca 608
 Db 545 CTGTGACGTGCTGGTGGAAAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCAACCA 604
 QY 609 ggaggaagacactgactgaattccctctgtgccaaacagctgctgaagggaagagacagag 668
 Db 605 GGAGGAAGACCTGACTGAATTTCTCTGTGCCAACACGCTGCTGAAGGGGAAGGACACGAG 664
 QY 669 ttgcttagcagagcgggtgtgtgtggaagagggggagacatagcctccctcgggaggaagaa 728
 Db 665 TTGCTTAGCAGAGCGGTGGTCTGGC-AGAAGGGGACATACCTTCCCTCGGTGGGAGGAAGAA 723
 QY 729 atccaagaagaagcgcagcgcgagtcgaagggtcctcctcagtcgagcagcagcagagagaa 788
 Db 724 ATCCAAGAAGAAGCGCAGCGGGTCAA-GGCTCTCCAGTGGCAGCAGCAGCAAGCAGAGAA 782

QY	789	ggaactggggg	799
Db	783	GGAACGCGG	793
RESULT	5		
BF785328			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
Qy	1063		
Db	1		
Qy	1123		
Db	61		
Qy	1183		
Db	121		
Qy	1243		
Db	181		
Qy	1303		
Db	241		

ORIGIN

Query Match 38.8%; Score 713; DB 10; Length 737;
Best Local Similarity 98.6%; Pred. No. 3e-122;
Matches 719; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1096 cagaaggagctctctttctacacagtcctccctcagctcagctcggggtcccccagcgc 1155
||| |
Db 6 CACGGCTCGCTCTCTTTCTACACAGTCCCTCTACAGAGCTCCGGGGTCCACCCAGCATC 65
Qy 1156 cccaggctgagatccaggtcctctgacatggaagctgaagacatgagggcacataagatgc 1215
||| |
Db 66 CCCAGGCTGAGATCCAGGCTCTGTACATGGAAGCTGAAGAGCATGAGGCACATAAGATGC 125
Qy 1216 tcaccagagccctctcagcgaaggaagactccgtgagcctcagcagcagcagcgcctgct 1275
||| |
Db 126 TCACGAGGCGCCCTTCACGAGGAGGAGCTCCGTCGAGCTCAGACGCGAGGCTGCCT 185
Qy 1276 ctctctccacacagcattctctctgctggtctcctgctgagtggaatttcgagaactt 1335
||| |
Db 186 CTTCTCTCCACCAAGCATTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245
Qy 1336 ccaggacaaatcgggtgtgacaaaggggtgagccagagccagagccagccagcaga 1395
||| |
Db 246 CCAGGACAAACTCGGGTGTGGCACAAGGGGTGGACGCCAGAGCCAGAGCCAGCCAGCA 305
Qy 1396 gactcagagagggcacctgacctaaccctcctggaaagcacaattcgagttcccgctgc 1455
||| |
Db 306 GACTCAGAGAGGGCACCCTGACCTTAACCCCTGCGGAAGCCCAATCTGCAGTTCCCGTGC 365
Qy 1456 caccactctctcagagagcctcctcctgctgcccagcctctctccagggctaccaga 1515
||| |
Db 366 CACCCACTCTCTCTGAGAGCGCTCATCTCTGCCACGCCCTCTCTCCAGGCTACCCAGA 425
Qy 1516 gtaaacacttttggccttctggttgggttctcctggtgctcagcctccagagtgctcc 1575
||| |
Db 426 GTAACACACTTTTGGCTTTTGGCTTTGGCTTTGGCTTTGGCTTTGGCTTTGGCTTTGG 485
Qy 1576 cctcctgactcttttgccttcttgcctcctcctcctcctcctcctcctcctcctcctc 1635
||| |
Db 486 CCTATCATCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCT 545
Qy 1636 attgaggcttaacactgctagtaggaggtgctgagggagcggcgggtggttggg 1695
||| |
Db 546 ATTGAGGCTTAACCTGTCACTTACTAGAGGTGCTGGAGGCGCGCGGGTGTGTATGGG 605
Qy 1696 gtaatacactgctgctcagcctcttaacactgagcccttaatacagttcctctg 1755
||| |
Db 606 GTAATCACTCACTGGCTCTCAGCCTTCTTAACACTGCAGCCCTTAATACAGTTCCCTCTG 665
Qy 1756 ttgtggtgactccacgccccac 1815
||| |
Db 666 TTGTTGTCACCTCCCGCCCCCAGACACACACACACACACACACACACACACACACACAC 725
Qy 1816 actgtaaaa 1824
||| |
Db 726 ACTGTACAA 734

RESULT 7

B1078951
LOCUS 602873203F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5005033 5',
DEFINITION mRNA sequence.
ACCESSION B1078951
VERSION B1078951.1 GI:14497281
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 836)

AUTHORS
TITLE
JOURNAL
COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11044 row: e column: 02
High quality sequence stop: 707.

FEATURES

source

1. 836
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5005033"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 187 a 265 c 210 g 174 t
ORIGIN

Query Match 38.7%; Score 711.4; DB 10; Length 836;
Best Local Similarity 95.2%; Pred. No. 5.9e-122;
Matches 768; Conservative 0; Mismatches 31; Indels 8; Gaps 3;

Qy 1041 gtgaaagatctgggtgctggagacggccaccccaagggaagagagggagagcaga 1100
Db 1 GTGAGAGATCTGGGGTCTGGGAGAGCGGCACCCCAAGGGAAGAGAGGAGCAGAA 60
Qy 1101 ggcagctctctttacacagtcctccctcagcagctccgggtccacccagcattccccc 1160
||| |
Db 61 GGCAGCTCTCTTTACACAGTCCCTCTACAGAGTCCGGGTCCAGCCAGCATCCCGAG 120
Qy 1161 gctagatccaggtcctcagatgaagctgaagcagcagcagcagcagcagcagcagcagc 1220
||| |
Db 121 GCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCATGAGGCACATAAAGATGCTCACC 180
Qy 1221 agcggcccttcagcaggaagagcctcgtgagcctcagcagcagcagcagcagcagcagc 1280
||| |
Db 181 AGCGCCCTTCACGAGGAGGACTCCGTGACGCTCAGCAGCCAGCCAGCCAGCCAGCTTCC 240
Qy 1281 ttccacaaagcattctctctgctggttctgctggtgctggtgctggtgctggtgctggtg 1340
||| |
Db 241 TTCCACCAAGCATTCCTCTCTGCTGGTCTGCTGGATGGTAAATTCGAGAACTTCCAGG 300
Qy 1341 acaactcgggtgtgacaaagggtgctgacgcacagccagccagccagccagccagccagccag 1400
||| |
Db 301 ACAACTCGGGTGTGGCACAAGGGGTGGACGCCAGCAGCCAGCCAGCCAGCAGACTG 360
Qy 1401 cagagagggcactgacctaaccctctggaagcctaactgacgttccctgctccccc 1460
||| |
Db 361 CAGAGGGGCACTTACCTAACCCCTCCCTGGGAAGCCCAATCTGCAGTTCCCGTGTCCACCC 420
Qy 1461 actcctcctgagagcagcctcagctgctgcccagccctctctccagggctaccagagtaaa 1520
||| |
Db 421 ACTCCTCTGAGAGCGGCTCATGCTCTGCCAGGCCCTTCTCCCAAGGGCTACCCAGAGTAAA 480
Qy 1521 cacttttggccttctggttctggttctggttctggttctggttctggttctggttctggtt 1580
||| |
Db 481 CACCTTTGGCTTTCTGGTTTGGTTCTGCTGGGTCTCTCATACAGCTCCAGAGTGTCCTCCCTCA 540
Qy 1581 tcgatctttttgctttgttcccccacaaatccca-ggggctggaaggccatcaccatcatgtg 1639


```
Qy 367 aagaggcttctggactacagcctgcacaaaggagagagactggcagcaaccgggttggccaag 426
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 360 AAGAGGGTCTGGACTACAGCTGCACAAGGAGGAGGACTGCGCAGCAACCGGTTGCCAAG 419
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 427 ggtatgtcggagaccttggagcgcctgcacaaacctagtcacaaagggttcaagtgtg 486
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 420 GGTATGTCGGAGACCTTTGAGACGCTGCACAAACCTTA-TCCCAAAGGGGTCAAGGTGGTG 478
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 487 atgatatccctctatgag-cgtgaaacagagacctcagcagagtggtgctgacctcaagaa 545
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 479 ATGGATATCCCTATGAGCCCTGTGAAGCAGACCTTCACAGAGGTGGCTGACCTCAAGAA 538
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 546 gcagtgtaactgctggtggaa-gagtttgaagaggtgattgagagactgggtacaggaacc 604
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 539 GCAGTGTACGCTGCTGGTGAAGGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACC 598
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 605 accaggaagaagacctgactgaattcctctgtgcacaaccacgctgctgaaggaagagaca 664
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 599 ACCAGGAGGAAGACCTGACTGAATTCCTGTGTCACACACGCTTGTGAAGGGAAGAGACA 658
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 665 cgagtgctcagcagacggtggtctgcaagaaggagacatagctccctcgagagga 724
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 659 CGAGTTGCTTACAGAGCGGTGGTCTGCGAAG-AGGGGGACATAGCCTTCCTGGGGGGGA 717
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 725 agaaatccaaagaagcgcgagtgcaagtgagggctcctccagtggcagcagcaa 779
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 718 AGAAATCCAAAGAAGACGCGACGGAATCAA-GGCTCTCCATGCGGAGCAGCA 771
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 11
BB498921
LOCUS BB498921 RIKEN full-length enriched, 0 day neonate kidney Mus
DEFINITION musculus cDNA clone D630017H14.3', mRNA sequence.
ACCESSION BB498921
VERSION BB498921.2 GI:16442133
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 645)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al., 2001)
Unpublished (2001)
On Jul 26, 2000 this sequence version replaced gi:9474466.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
```

10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001).
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES
 Location/Qualifiers
 1..645
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="D630017H14"
 /clone_lib="RIKEN full-length enriched, 0 day neonate
 kidney"
 /tissue_type="kidney"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site_1: Sali; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGCGCGCCAACTCGAGTTTTTTTTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGAGAGAGATTCTCGAGTTAAATAAATTAATCCGCCCCCCCCCC 3']. cDNA
 was cleaved with BamHI and XhoI. Vector: a modified
 pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 137 a 209 c 150 g 148 t 1 others
ORIGIN
 Query Match 34.7%; Score 638.2; DB 9; Length 645;
 Best Local Similarity 99.4%; Pred. No. 2.2e-108;
 Matches 640; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1176 cctgacatggaagctgaagcagcatgaggcacaataagtcgtccaccagcgcccccttcagc 1235
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 CCTAATCGAAGCTAAAGAGCATGAGGCACATAAGATGCTCACCAGCGCCCCCTTCAGC 60
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1236 caggaagaagcctcgtgagcctcagcagcaggcgtcctctcttccacaagcattc 1295
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 CAGGAAGGACTCCGTGCGAGCTCAGACGCCAGGCGTCTCTCTTCTCCACCAAGCATTC 120
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1296 tcttctgctgcttctgtcggatggtataattcgagaacttccagacaaactcgggtgtg 1355
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 TCTTCTGTGTCCTCTGTTCGGAAGGTAATTCGAGAACTTCAGAGACAAACTCGSGGTGTG 180
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1356 gcacaaagggttgagccagcagcagcagccagccagcagactgcagagaggccactg 1415
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 181 GCACAAAGGGGTGTGACGCCAGAGCCAGCCAGAGACTGCAGAGAGGGCACCTG 240
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1416 acctaaacccctggaagaacaaactcgtcagttccgttccaccactcctcctgagcag 1475
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 241 ACCTAACCCCTCGAAAGCCAACTCTGCAGTTCCCGTGTCCACCCACATCTCTCTGAGGAC 300
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 1476 gcctcatgctctgccacgacctcttcccaggggtaccagagtaaacaccttttggccttt 1535
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db	301	CGCTCATGCTCTGCCACGCCCTTCTCCAGGGCTACCAAGAGTAACACACTTTTGGCCTTT	360
Qy	1536	cggttgggttctctgggtctcatcagcctcagagtgccccctcatcagatctttttgccc	1595
Db	361	CGGTTTGGTTCCTGGGTCTCATCAGCCTCCAGAGTGTCCTCCCTCATCATCTTTTGGCC	420
Qy	1596	tttgtcccccataccccagggggctggaaggccatcacatcattggaggcttaacctgtca	1655
Db	421	TTTCTCCCCCAATCCAGGGGCTGGAAGGCCATCACCATCATTTGAGGCTTAACCTGTCA	480
Qy	1656	gttactaggaggtactgggaagcgcggggggtgtgttggggtaatacactcactcggctctc	1715
Db	481	GTTTACTAGGAGGTGCTGGGAGCGGCCGGGGTGGTTGGGGTAATCACTACTGGGCTCTC	540
Qy	1716	agccttctaaactgcagcccttaatacacagtctctctgtgtggtgaactccacagccc	1775
Db	541	AGCCTTCTAACACTGCAGGCCCTTAATACAGTTCTCTCTGTGTGGTGACTCCACAGCCC	600
Qy	1776	ccacacacacacataaaattatttcgatgctgtttcataaactg	1819
Db	601	CCACACACACACCAATAAAATATTTCGATGCTGTTCATAACTG	644
RESULT	12		
BB637089			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

QY 488 ttgatatcccttatgagctgtggaagagacctcagcagagtgctgacctcaagaagc 547
 |||||
 Db 481 TGGATATCCCTTATGAGTGTGGAACGAGACCTCAGCAGAGTGCTGACCTCAAGAAGC 540
 |||||
 QY 548 agtgagcgtgctggtggaagagtttgaagagtgatgagagctgggtacaggaaccacc 607
 |||||
 Db 541 AGTGACGTGCTGCTGGTGAAGAGTTTGAAGAGTGATTGAGAGCTGGTCCAGGAACCACC 600
 |||||
 QY 608 aggaggaagacctgactgaattcctctgtgccaacacagtg 648
 |||||
 Db 601 AGGAGGAAAACCTGACTGAATTCCTCTGTGTGCCAACCAAGCTG 641
 |||||

RESULT 13

BB190512 644 bp mRNA linear EST 19-OCT-2001
 BB190512 RIKEN full-length enriched, adult male spinal cord Mus
 musculus cDNA clone A330062B19 3', mRNA sequence.

ACCESSION BB190512
 VERSION BB190512.2 GI:16270724

KEYWORDS

SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

AUTHORS 1 (bases 1 to 644)
 Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
 M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Sakai,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
 D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

TITLE

JOURNAL

COMMENT

On Jun 30, 2000 this sequence version replaced gi:8851131.
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,I., Aizawa,
 K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
 Hayashizaki,Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES

source

Location/Qualifiers

1..644
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A330062B19"
 /clone_lib="RIKEN full-length enriched, adult male spinal
 cord"
 /sex="male"
 /tissue_type="spinal cord"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Site 1: Sali; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 459.0. Second
 strand cDNA was prepared with the primer adaptor of
 sequence [5' GAGAGAGAGATCTCGAGTAAATAATATCCCTCCCTCC
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 FLC I."

BASE COUNT 134 a 209 c 151 g 150 t

ORIGIN

Query Match 34.4%; Score 632.8; DB 9; Length 644;
 Best Local Similarity 99.7%; Pred. No. 2.2e-107;
 Matches 634; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1185 gaagctgaagcagcatgagcgacataagatgctcaccagcgcccccttcagccagaagga 1244
 |||||
 Db 9 GAAGCTGAAGAGCATGAGGCACATAAGATGCTCACCAGCGCCCCCTCAGCAGGAAGA 68
 |||||
 QY 1245 ctccgtgagcctcagcagcgccctcctctctccaccagcattctctctct 1304
 |||||
 Db 69 CTCCTGTGAGCCTCAGCAGCTAGGCTGCTCTCTCTCCACCAAGCAATCTCTTCTGCT 128
 |||||
 QY 1305 ggtccttgcgagtggttaattcgagaacttcacagacaaactcgggtgtggcacaagg 1364
 |||||
 Db 129 GGTCTCTTGTGGATGTTAAATTCGAGAACTTCCAGGACAACTCGGGTGTGGCACAAAG 188
 |||||
 QY 1365 ggctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1424
 |||||
 Db 189 GGCTGGAGCCAGAGCCAGAGCCAGCCAGAGACTGCGAGAGAGGCGCACTGACCTTAACCC 248
 |||||
 QY 1425 ccttgagaagccaatctcagttcccggtgtccaccactctcctgagagcagcctcagtc 1484
 |||||
 Db 249 CCTTGGAAGCCCAATCTCGAGTTCCTCGGTGTCACCCACTCTCTCTCTGAGAGCGCTCATGC 308
 |||||
 QY 1485 tctgcccagccctctccacggggtaccagagtaaacacaccttttggcctttgtggttgg 1544
 |||||
 Db 309 TCTGCCAGCCCTTCTCCAGGGCTACCAGAGTAAACACCTTTTGGCCTTTTCGGTTTGT 368
 |||||
 QY 1545 tcttggtctctcagcagcctcagagtgctccctcagctcagctcagctcagctcagctcagc 1604
 |||||
 Db 369 TCTTGGGTCTCTCATCAGCCTCCAGAGTGTCCCTCATCATGATCTTTTGGCCTTTGTGCC 428
 |||||
 QY 1605 caatcccaggggtggaagccatcacatcatgtgagagcttaacctgtcagttactagg 1664
 |||||
 Db 429 CAATCCCAGGGCTGGAAGGCCATCACCATCATTTGGAGGCTTAACCTGTCACTTACTAGG 488
 |||||
 QY 1665 aggtgctgggagcgcccggtgtgtgtgggttaatacactcagctggctcagctcagctcagc 1724
 |||||
 Db 489 AGGTGCTGGGAGCGCCCGGGTGTGGTTGGGGTAACTCACTCACTGCTGCTCAGCCTCTCTA 548
 |||||

Db	193	TTTCAGAGCTTTGGGACACGACAGCGCAGCGCAGCTCCACCAAGACAGCTGCTGT	252
Qy	981	ccagcatcaggctccctcccttgctgctgcccccttctcccttgaaacaagcaag	1040
Db	253	CCACCATCAGGTCTCCTTGGCTGTGCCCCCTTCCCTTGAACAAACAGCAAG	312
Qy	1041	gtgaagaagatctggggctgctgggagacgcccccaaggaagagagagagcagaa	1100
Db	313	GTGAAGAAGATCTGGGTGCTGGGACAGCGCACCCCAAGGAAGAGGAGGAGCAGAA	372
Qy	1101	ggcagctctctttctacacagtcctccctcagagctcgggggtccacccagatcccccag	1160
Db	373	GGCAGCTCTCTTTCTACACAGTCCCTCTCAGAGCTCCGGGTCCACCCAGATCCCCAG	432
Qy	1161	gctgagatccaggctctgacatgaagctgaagcatgagggcacataagatgctcacc	1220
Db	433	GCTGAGATCCAGGCTCCTGACATGGAAGCTCAAGAGCATGAGGCACATAAGATGCTCACC	492
Qy	1221	agcccccttcagccagggaaggaactcgtgagcctcagcagcaggctgcctctccc	1280
Db	493	AGCCCCCTTCAGCCAGGAAGGACTCCGTGCAGCCTCAGCAGCAGCGCTGCCCTCTTCC	552
Qy	1281	ttccaccaagcatctctctgctggtcccttgctcgatgtagtaaatcagaaactccagg	1340
Db	553	TTCCACCAAGCATCTCTTCTGCTGGTTCCTGTCGGATGCTACATCGAGAACTTCAGG	612
Qy	1341	acaaactcgggtggtggcacaaaaggg-ctggagccagagcc	1381
Db	613	ACAACTCGGGTGTGGCACAAAGGGGCTGGACCCAGAGCC	654

RESULT 15

BB385488

LOCUS

BB385488 RIKEN full-length enriched, 0 day neonate cerebellum Mus musculus cDNA clone C230039G09 3', mRNA sequence.

ACCESSION

BB385488

VERSION

BB385488.2

KEYWORDS

GI:16408338

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 651)

AUTHORS

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Sasaki,O., Shikata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

COMMENT

Unpublished (2001)

On Jul 13, 2000 this sequence version replaced gi:9108269.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shikata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multipillarary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamana, K. I., Aizawa, K., Fukuda, S., Harada, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.riken.go.jp/>) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES

Location/Qualifiers
 1. .651
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="C230039G09"
 /clone_lib="RIKEN full-length enriched, 0 day neonate cerebellum"
 /tissue_type="cerebellum"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTIVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATTAATCCCTCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified plasmid KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 135 a 207 c 154 g 155 t
 ORIGIN

Query Match 34.1%; Score 626.4; DB 9; Length 651;
 Best Local Similarity 99.1%; Pred. NO. 3.3e-106;
 Matches 630; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1185 gaagctgaagacatgagggcacataagatgtccaccagcgcccccttcagccaggaaagga 1244
 |||||
 DB 16 GAAGTTGAAGAGCATGAGGCACATAGATGCTCACAGCGCCCCCTTCAGCAGGGAAGGA 75
 |||||

QY 1245 etccgtgcagctcagcagcagcgcctcctctcttcctccaccagcattctctctgt 1304
 |||||
 DB 76 CTCCTGTCAGCTTCAGCAGCAGGCTGCCCTCTTCTTCCACCAAGCATTTTCTGCT 135
 |||||

QY 1305 ggtccttgcggatggttaattcgagaacttcaggacaaactcgggtgtggcacaaagg 1364
 |||||
 DB 136 GGTCTTGTGGATGGTAAATTCGAGAACTTCACAGGACAAACTCGGGTGTGGCACAAAGG 195
 |||||

QY 1365 ggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1424
 |||||
 DB 196 GGCTGGAGCCAGAGCCAGAGCCAGAGACTGCAGAGAGGGCAGCTGACCTAACCC 255
 |||||

QY 1425 cctcgtgaagccaattctcagttccctgtccaccactcctcctcgtgagcagcctcatgc 1484
 |||||
 DB 256 CCTCGAAGCAATCTGCAGTTCCTCGTGTCCACCATCTCTCTTGAGAGCGCTCATGC 315
 |||||

QY 1485 tctgccagggcccttctccaggggtaccagagtaaacacccttttggcctttcggtttgt 1544
 |||||
 DB 316 TCTGCCAGCCCTTCTCCAGGGCTACCAGAGTAACACACCTTTTGGCCTTTTCGGTTGGT 375
 |||||

QY 1545 tcttggttccctcatcagcctccagagtggtccctcatcgatctttttgctttgtccccc 1604
 |||||
 DB 376 TCTTGGGTCTCATCAGCCTCCAGAGTGTCCTCATCATGATCTTTTGGCCTTTGTCCCC 435
 |||||

QY 1605 caatccccggggtggaaggccatcacccatcatttgaggcttaacctgtcagttactagg 1664
 |||||
 DB 436 CAATCCCAGGGGCTGGAAGGCCATCACCATCATTTGGAGGCTTAACCTGTCACTTACTAGG 495
 |||||

QY 1665 aggtctctggagcgccccgggtgtgttggttggttggttggttggttggttggttggt 1724
 |||||
 DB 496 AGGTCTCTGGAGCGCCCGGGTGGTTGGTGGGTAATCACTCACTGGCTCTCAGCCTTCTA 555
 |||||

QY 1725 acactgcagcccccttaataacagttctctctgtgtgtgtgtgtgtgtgtgtgtgtgt 1784
 |||||
 DB 556 ACACGTGCAGCCCTTAATACAGTTCCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 615
 |||||

QY 1785 caccataaaattatttcgatgctgtgtttcataaactgt 1820
 |||||
 DB 616 CACCATATAAATTTATTTCGATGCTGTTTTCATAAATGT 651
 |||||

Search completed: May 23, 2002, 23:05:57
 Job time: 7513 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:18:10 ; Search time 121.39 Seconds
(without alignments)
3721.229 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgc.....taaaaaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6ptodata/1/ina/PCITUS_COMB.seq.*
6: /cgn2_6ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	160.2	8.7	744	4	Sequence 3, Appli
2	160.2	8.7	1512	4	Sequence 1, Appli
3	75.2	4.1	7218	1	Sequence 14, Appl
4	56.2	3.1	7218	1	Sequence 14, Appl
5	55.4	3.0	1931	2	Sequence 2, Appli
6	48.6	2.6	3489	2	Sequence 1, Appli
7	48.6	2.6	32207	2	Sequence 20, Appl
8	48.6	2.6	32207	4	Sequence 20, Appl
9	48.6	2.6	32207	4	Sequence 20, Appl
10	46.8	2.5	289	4	Sequence 17, Appl
11	46.8	2.5	289	4	Sequence 17, Appl
12	44.6	2.4	3319	1	Sequence 2, Appli
13	44.6	2.4	3319	1	Sequence 2, Appli
14	44.6	2.4	3319	1	Sequence 4, Appli
15	44.6	2.4	3319	5	Sequence 4, Appli
16	44.4	2.4	2580	3	Sequence 2, Appli
17	44.4	2.4	2580	4	Sequence 2, Appli
18	44.4	2.4	5452	2	Sequence 1, Appli
19	44.4	2.4	9600	4	Sequence 1, Appli
20	44.4	2.4	10596	1	Sequence 15, Appl
21	44.4	2.4	10596	1	Sequence 15, Appl
22	44.4	2.4	10596	1	Sequence 15, Appl
23	44.4	2.4	10596	1	Sequence 15, Appl
24	44.4	2.4	10596	2	Sequence 15, Appl
25	44.4	2.4	10596	5	Sequence 15, Appl
26	39.2	2.1	24979	2	Sequence 3, Appli
27	39.2	2.1	24979	3	Sequence 3, Appli

c	28	39.2	2.1	24979	5	PCT-US93-03985-3	Sequence 3, Appli
	29	38.2	2.1	1575	3	US-08-478-507-11	Sequence 11, Appl
	30	38.2	2.1	1575	4	US-09-128-275A-11	Sequence 11, Appl
	31	38.2	2.1	16442	3	US-08-781-891-208	Sequence 208, App
c	32	38.2	2.1	53526	3	US-08-658-136-2	Sequence 2, Appli
	33	38.2	2.1	53577	3	US-08-658-136-1	Sequence 1, Appli
c	34	37.6	2.0	16442	3	US-08-781-891-208	Sequence 208, App
	35	37.2	2.0	4112	1	US-08-340-203A-2	Sequence 2, Appli
c	36	37.2	2.0	4112	2	US-08-452-567-2	Sequence 2, Appli
	37	37.2	2.0	4112	2	US-08-452-567-2	Sequence 2, Appli
c	38	37.2	2.0	4112	3	US-09-085-407-2	Sequence 2, Appli
	39	37.2	2.0	4616	1	US-08-340-203A-1	Sequence 1, Appli
c	40	37.2	2.0	4616	2	US-08-452-567-1	Sequence 1, Appli
	41	37.2	2.0	4616	2	US-08-452-567-1	Sequence 1, Appli
c	42	37.2	2.0	4616	3	US-09-085-407-1	Sequence 1, Appli
	43	37.2	2.0	51259	3	US-08-781-891-209	Sequence 209, App
c	44	37	2.0	1433	2	US-08-666-392A-1	Sequence 1, Appli
	45	37	2.0	1433	4	US-09-199-926-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-163-285-3
; Sequence 3, Application US/09163285
; Patent No. 6204013
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/163,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/090,398
; FILING DATE: June 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-049
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..744
; US-09-163-285-3

Query Match 8.7%; Score 160.2; DB 4; Length 744;
Best Local Similarity 57.3%; Pred. No. 8.5e-32;

[illegible]

```

2
RESULT
US-09-163-285-1
; Sequence 1, Application US/09163285
; Patent No. 6204013
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/090,398
; FILING DATE: June 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.

```

```

; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 125..868
; US-09-163-285-1

```

	Query Match	8.7%	Score 160.2	DB 4	Length 1512
	Best Local Similarity	57.3%	Pred. No. 1.le-31		
	Matches 310	Conservative	0	Mismatches 228	Indels 3
					Gaps
Qy	162	ggctgagagaccgactgggtgcgattgccacgaatcgaaagtgtcgaagtattgttc	221		
Db	199	GGAGGAGGACGATGACACAGAAGCGTTGCCACGAATGCGAAGTGTGAAGCTGCTGAG	258		
Qy	222	tgtgaactgaatcgcttttgagaaacggaaagaccacgaaggaagtgtatgacacgg	281		
Db	259	CACAGAGCTACAGCGGAACGTGAGTCCACCGTGCATCTCGAGAGGTGCTGGAGCTGG	318		
Qy	282	ctatggcatcctggacggggaagggtctctggagtcaag --- tacaccaagtctggacttacg	338		
Db	319	GCAGGTGCTGGATACAGGCAAGAGGAAGACACAGTCGCTTACAGCGTTTCAGAGACAAG	378		
Qy	339	gttaattgaagtcactgagaccatttgcgaagggctcttgacctacagcctgcacaagga	398		
Db	379	GCTGGAGAGGCCTTACAGAAATTATGTGAGCGGATCCTGGACTATAGTGTTCACGCTGA	438		
Qy	399	gaggactggcagcaacccgggttgcgaagggtatgtcggagacccttggagacgctgcacaa	458		
Db	439	CGCGAAGGGTTCACGTAGATATGCCAAGGGTTCAGAGTCAGACCATGCAACACTGAAGG	498		
Qy	459	ctagtccacaaaggggtcgaaggtggttgatggatatccctctatgagctgtggaaacgagac	518		
Db	499	CCTAGTCAGAAAGGGGTGAAGGTGGATCTGGGATATCCCTCTGGAGCTTTGGGATAGCC	558		
Qy	519	ctcgcgagagtggtgacctcaagaagcagtgatgacgtctgctggtggaagagttaaaga	578		
Db	559	CAGCGTGAGGTGCACATACCTCAAGAGACGAGTGTGAGACCATGTGTGGAGGAGTTGAAG	618		
Qy	579	ggtgattgagactggttacaggaaaccacggaggaagcactgactgaattctctctgtgc	638		
Db	619	CATTGTGGAGACTGGTACTTCCACCATCAGGAGCAGCCCTACAAAAATTTCTCTGTGA	678		
Qy	639	caaccacgtgctgaagggaagaaacacagttacctcagcagacgggtggcttggcgaagaa	698		
Db	679	AGGTCACTGTCTCCACGTGCTGAACCTGCATGCTTACAGGAACCTTGGACTGGAAAGGA	738		
Qy	699	g	699		
Db	739	g	739		

RESULT 3
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHAEFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52

```

; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
;
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgpt-Fls
; US-08-232-463-14

```



```

Qy  746  gcggagtcagggtctctccagtgcagcagcaagcaagaaagaaactgggggcttg  805
Db  2541  AGAGCAGGAGGTGGAAGAGAGCAGGAGCAGGAGCAGGAGCAGCAGC-AGGAAATTAGAGGAGGTGG  2599
Qy  806  ggagagatgccaacgccaggaggaggagggtgtgtcagaag  846
Db  2600  AGGAGCAAGAGCAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAG  2640

RESULT      7
US-08-770-379-20/c
: Sequence 20, Application US/08770379
: Patent No. 5849564
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Bohenzky, Roy A.
: APPLICANT: Russo, James J.
: APPLICANT: Edelman, Isidore S.
: APPLICANT: Moore, Patrick S.
: TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
: TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/770,379
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 52342
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 32207 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

```

Query Match	2.6%	Score 48.6;	DB 2;	Length 32207;
Best Local Similarity	48.4%;	Pred. No. 0.012;		
Matches 194;	Conservative 0;	Mismatches 204;	Indels 3;	Gaps 2;
QY 446	agacgctgcacaacctagttcccaaaaggggtccaaggtggtgatgatatcccttatgacg	505		
Db 19754	AGCAGCAGGATGACACAGGACGACGAGGAGCAGGAGCAGGAGGACGAGGACGG	19695		
QY 506	tgttggaacgagacctcagcagaggttggtcgtacctcaagaagcagtgtagctgctggtg	565		
Db 19694	AGTTAGAGGAGCAGGACGAGGAGTTAGAGGATCAGGAGCAGGAGCTTAGAGCAGCAGGAGC	19635		
QY 566	aagagtttgaagaggtgatgtaggacttggtacaggaaccaccagaggaagacctgactg	625		
Db 19634	AGCAGTTAGAGGAGCAGGACGACGAGTGTAGAGGACGAGGACGAGGTTAGAGGACGAGG	19575		
QY 626	aattcctctgtgcacaaccacgtgctgaagggaaggacacagagtgtgctagcagagcggt	685		

```

Db   19574 AGCAGGACTTTAGAGGACGACGAGGAGGTTAGAGGACC--AGGAGCAGGAGTTCAGAGCA    19517
QY     686 ggtctgccaagaaggggcacatagacctccctccttggaaggaaatacccccaagaagcgaca    745
Db   19516 GCAGAGGACGAGGAGTTAGAGGAGCACGAGGACGAGCTTAGAGGACGAGCGTGGAAGAACA    19457
QY     746 gcggagtcaaagggcttcccacttcagttgccagcacgaagcagaggaaggaagctggggggcctcg    805
Db   19456 AGAGCAGGAGTGGAAGAAGCAAAGACGACGAGGACGAGGACGAGGAAAGAGC-AGCAATTAGAGGAGGTGG    19398
QY     806 gggaggattaccaagcaccgagtagagagggagggctgtcgcagaag      846
Db   19397 AGGAGCAAGACGAGGACGAGGACGAGGACGAGGAGGAGGAGGACGAGGAG    19357


RESULT           8
US-08-757-669A-20/C
; Sequence 20 Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEX/FAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20
```

[illegible]


```
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17
```

```
Query Match          2.5%; Score 46.8; DB 4; Length 289;
Best Local Similarity 5.3%; Pred. No. 0.0054;
Matches 11; Conservative 101; Mismatches 96; Indels 0; Gaps 0;

Qy 640 aaccacgtgctgaagggaagacacagattgacctagcagagcggtggtctggcgaagaag 699
      ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 45 rarcrarurgrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 104

Qy 700 ggggacatagctccctcctggggggaagaataccaaagaagcgcagcggaagtcaagggc 759
      ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 105 rnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 164

Qy 760 tcctccatggtcagcagcaacagcaggaagaaactggggcctgggggagagatgccaac 819
      ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 165 rnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 224

Qy 820 gccagaggaggaggggtgtgcagaag 847
      ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 225 rcrrgrgrgrarararurururgr 252
```

```
RESULT 12
US-08-006-676B-2
; Sequence 2, Application US/08006676B
; Patent No. 5411865
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven
; TITLE OF INVENTION: Diagnosis of Leishmaniasis
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeffrey B. Oster
; STREET: 8339 SE 57th Street
; CITY: Mercer Island
; STATE: Washington
; COUNTRY: USA
; ZIP: 98040-4906
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORD for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/006,676B
; FILING DATE: 15-JAN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oster, Jeffrey B.
; REGISTRATION NUMBER: 32,585
; REFERENCE/DOCKET NUMBER: REED-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 232 7845
```

```
; TELEFAX: (206) 236 0205
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Leishmania chagasi
US-08-006-676B-2

Query Match          2.4%; Score 44.6; DB 1; Length 3319;
Best Local Similarity 46.1%; Pred. No. 0.051;
Matches 149; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

Qy 419 ttgccaagggtatgtcggagacctttgagacgtgtgcacaacctagtcacacaagggggtca 478
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1629 TTGTGAATGTTGGCGAGGTGAACGAGGACCCCGCCGCGGATCCCGGAGCTGGAGG 1688

Qy 479 aggtggtgatgatccctatgagctgtggaacgagacctcagcagaggtggtgtacc 538
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1689 AGCAGATGAGGACATGCGGACGCGATGGCTGGCGGCGACCCCGCGTACGTGTCTGAGC 1748

Qy 539 tcaagaagcagtgtaactgtgtgtggaagagtttgaagaggtgattgaggactggtaca 598
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1749 TGAAGAAGAAGTTGCGCTGCTGGAGTCGGAGGCGCAGAACGCTCGCGCGGACCTGCAGG 1808

Qy 599 ggaacccaccagagggaagacctgactgaattcctctgtccaaccacgtgctgaaggaa 658
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1809 CGCTGAGAGGAGCGGGAGACACACCAGGTGCAGGAGCGGCTGCTGCCGCGGAGGAGG 1868

Qy 659 aggcacacagtgctcctagcagagcgtgtgtctggaagagggggacatagcctccctgg 718
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1869 CGGAGAAGAGCGAGCTGGAGTCGCTGGCGCTGCGTGCAGGAGGAGATGACCGCGACTC 1928

Qy 719 gaggaagaataccaaagaag 741
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1929 GACGCGGCGGACAGATGCAG 1951
```

```
RESULT 13
US-08-282-845-1
; Sequence 1, Application US/08282845
; Patent No. 5719263
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: A 230kd Antigen Present in Leishmania
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple Macintosh Operating System 7.1
; SOFTWARE: Microsoft Word for Macintosh 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,845
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/006,676
; FILING DATE: JANUARY 15, 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia Anne
```

REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 5004-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: rk39
FEATURE:
NAME/KEY: CDS
LOCATION: 455..3319
US-08-282-845-1

Query Match 2.4%; Score 44.6; DB 1; Length 3319;
Best Local Similarity 46.1%; Pred. No. 0.051;
Matches 149; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 419 ttgccaagggtatgtcggagacacctttgagacgtgcacaaacctagtccacaaagggtca 478
DB 1629 TTGTGAATGTTCCGAGGTGAACGAGGACCCGCGGCACCGCGGATCCGCGAGCTGGAGG 1688
QY 479 agtgtgtgatgatccccctatgactgtggaacgagacctcagcagaggttgctgacc 538
DB 1689 AGCAGATGGAGGACATGCGCGAGGCGATGGCTGGCGGCACCCCGCTACGTGTCTGAGC 1748
QY 539 tcaagaagcagtgtagctgtcgtgtggaagagtttgaagagtgattgagagactgtaca 598
DB 1749 TGAAGAAGAGCTTGGCTGTCTGGAGTCGAGCGCGCAGAAAGCGTGGCGGACCTGCAGG 1808
QY 599 ggaaccaccagagagaagacctgaattcctctgtgccaaaccacgtgtcgaaggaa 658
DB 1809 CGCTGGAGAGGAGCGGGAGCACACACAGGTGCAGAGCGGCTGTGCGCGCGAGGAGG 1868
QY 659 agcacacagtgctacagagcgtgtcgtgcaagaaggagacatagcctccctgg 718
DB 1869 CGGAAGAAGAGCGAGCTGGAGTCGCTGCGGCTGCGCTGCAGGAGGAGATGACCGGACTC 1928
QY 719 gagggagaagaatcccaagaagaag 741
DB 1929 GACGGCAGCGGACAAAGATGCAG 1951

RESULT 14
PCT-US94-414A-4
Sequence 4, Application US/08428414A
Patent No. 5912166
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TITLE OF INVENTION: LEISHMANIASIS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-APR-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlecak, Ann T.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121.407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-428-414A-4

Query Match 2.4%; Score 44.6; DB 2; Length 3319;
Best Local Similarity 46.1%; Pred. No. 0.051;
Matches 149; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 419 ttgccaagggtatgtcggagacacctttgagacgtgcacaaacctagtccacaaagggtca 478
DB 1629 TTGTGAATGTTCCGAGGTGAACGAGGACCCGCGGCACCGCGGATCCGCGAGCTGGAGG 1688
QY 479 agtgtgtgatgatccccctatgactgtggaacgagacctcagcagaggttgctgacc 538
DB 1689 AGCAGATGGAGGACATGCGCGAGGCGATGGCTGGCGGCACCCCGCTACGTGTCTGAGC 1748
QY 539 tcaagaagcagtgtagctgtcgtgtggaagagtttgaagagtgattgagagactgtaca 598
DB 1749 TGAAGAAGAGCTTGGCTGTCTGGAGTCGAGCGCGCAGAAAGCGTGGCGGACCTGCAGG 1808
QY 599 ggaaccaccagagagaagacctgaattcctctgtgccaaaccacgtgtcgaaggaa 658
DB 1809 CGCTGGAGAGGAGCGGGAGCACACACAGGTGCAGAGCGGCTGTGCGCGCGAGGAGG 1868
QY 659 agcacacagtgctacagagcgtgtcgtgcaagaaggagacatagcctccctgg 718
DB 1869 CGGAAGAAGAGCGAGCTGGAGTCGCTGCGGCTGCGCTGCAGGAGGAGATGACCGGACTC 1928
QY 719 gagggagaagaatcccaagaagaag 741
DB 1929 GACGGCAGCGGACAAAGATGCAG 1951

RESULT 15
PCT-US94-00324-2
Sequence 2, Application PC/TUS9400324
GENERAL INFORMATION:
APPLICANT: Reed, Steven
TITLE OF INVENTION: Diagnosis of Leishmaniasis
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1
SOFTWARE: Microsoft Word, version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00324
FILING DATE:
CLASSIFICATION:

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:23:11 ; Search time 525.6 Seconds
(without alignments)
6007.241 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgcc.....taaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match	Length	ID		
1	1839	100.0	1839	22	AAS05278	Murine trinucleotid
2	1725.2	93.8	1848	22	AAS05286	DNA sequence of an
3	691.6	37.6	1675	22	AAS21311	Human CDNA sequenc
4	665.8	36.2	1362	22	AAS05279	Human trinucleotid
5	446	24.3	1657	20	AAS04377	Human secreted pro
6	412.2	22.4	580	22	AAS25051	Human ovarian PCR-
7	412.2	22.4	580	22	AAH83696	Human ovarian tumo
8	224.4	12.2	274	21	AAA42779	Human secreted exp
9	160.2	8.7	744	21	AAZ49678	Human myocardium s

10	160.2	8.7	746	22	AAF94104	Primer specific fo
11	160.2	8.7	747	22	AAC90704	Human secretory pr
12	160.2	8.7	1419	22	AAF93914	Human CDNA encodin
13	160.2	8.7	1457	20	AAF78923	Human zcalc-1 DNA
14	160.2	8.7	1462	21	AAA16621	Human secreted pro
15	160.2	8.7	1470	21	AAAG6349	CDNA encoding a no
16	160.2	8.7	1470	21	AAC58116	Human PRO4354 nucl
17	160.2	8.7	1470	22	AAS46164	Human DNA encoding
18	160.2	8.7	1512	21	AAZ49677	Human myocardiolum s
19	156.6	8.5	1658	21	AAZ98029	Human secreted pro
20	156.6	8.5	1658	22	AAZ11642	Human secreted pro
21	151.4	8.2	750	23	ABL16485	Drosophila melanog
22	151.4	8.2	2750	23	ABL16484	Drosophila melanog
23	151.4	8.2	3274	23	ABL16482	Drosophila melanog
24	142.4	7.7	189	16	AAT23979	Human gene signatu
25	137	7.4	1623	20	AAH88064	Human zcalc-1 DNA
26	123.6	6.7	1306	20	AAZ24902	Human secreted pro
27	123.6	6.7	1312	22	AAZ05462	Human secreted pro
28	117.8	6.4	1315	22	AAZ05464	Human secreted pro
29	107	5.8	107	22	AAS05289	DNA sequence of de
30	98.6	5.4	293	20	AAZ32038	Human METH2 relate
31	98.6	5.4	293	22	AAC90095	HTOFC34RA cDNA clo
32	69.4	3.8	4069	23	ABL11794	Drosophila melanog
33	67.6	3.7	501	22	AAH29134	Drosophila melanog
34	67.6	3.7	510	22	AAH29134	Umbilical vein end
35	66	3.6	471	22	AAS05284	T243 gene homolog
36	57.4	3.1	10732	21	AAAL0594	Gene encoding a su
37	51	2.8	443	20	AAH88063	Human zcalc-1 DNA
38	49.4	2.7	1416	23	AAS71051	DNA encoding novel
39	49.4	2.7	4720	24	AAH94782	Human DNA sequence
40	49.4	2.7	5240	22	AAK52624	Human polynucleoti
41	49.4	2.7	5360	22	AAK51640	Human polynucleoti
42	48.6	2.6	3489	21	AAA30290	Kaposi's sarcoma-a
43	48.6	2.6	3489	22	AAH82901	Nucleotide sequenc
44	48.6	2.6	32207	20	AAH73805	KSHV LUR DNA (nucl
45	48.6	2.6	137507	19	AAV19941	KSHV long unique c

ALIGNMENTS

RESULT 1	
AAS05278	
ID	AAS05278 standard; cDNA; 1839 BP.
XX	
AC	AAS05278;
XX	
DT	07-SEP-2001 (first entry)
XX	
DE	Murine trinucleotide repeat protein (TRP) cDNA sequence.
XX	
KW	Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW	transgenic animal; knockout mouse; triplet repeat expansion;
KW	fragile X syndrome; Huntington's disease; mouse; murine; ss.
XX	
OS	Mus sp.
XX	
FH	Key
CD	Location/Qualifiers
FT	55..885
FT	/*tag= a
FT	/product= "TRP"
XX	
PN	WO200130798-A1.
XX	
PD	03-MAY-2001.
XX	
PF	26-OCT-2000; 2000WO-US29382.
XX	
PR	26-OCT-1999; 99US-0161488.
XX	
PA	(DELT-) DELTAGEN INC.
XX	
PI	Klein R, Matthews W, Moore M, Allen KD;

|||||
Db 1621 aagccatccattgaggcttaacctgactagtagaggtgctggagcgc 1680
Qy 1681 cggggttggtgggtaatacactgctgctcagcctcttaacactgcagccctta 1740
Db 1681 cggggttggtgggtaatacactgctgctcagcctcttaacactgcagccctta 1740
Qy 1741 atacagttctctctgtgtgactccacgcgcgcacacacacataaaattattt 1800
Db 1741 atacagttctctctgtgtgactccacgcgcgcacacacacataaaattattt 1800
Qy 1801 cgatgctgtttcataactgtataaaaaa 1839
Db 1801 cgatgctgtttcataactgtataaaaaa 1839
RESULT 2
AAS05286
ID AAS05286 standard; DNA: 1848 BP.
XX AAS05286;
AC AAS05286;
DT 07-SEP-2001 (first entry)
XX DNA sequence of an expanded T243 gene.
XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; mutant; ds.
XX Mus sp.
OS Synthetic.
XX Key Location/Qualifiers
FT CDS 55..894
FT FT /*tag= a
FT FT /product= "Expanded TRP"
XX WO200130798-A1.
XX 03-MAY-2001.
XX 26-OCT-2000; 2000WO-US29382.
XX 26-OCT-1999; 99US-0161488.
XX (DELTA-) DELFAGEN INC.
XX Klein R, Matthews W, Moore M, Allen KD;
XX WPI: 2001-300473/31.
XX P-PSDB; AAU02500.
XX Novel transgenic animals useful as animal model for characterization of
XX function of a gene encoding trinucleotide repeat proteins (TRPs),
XX contains heterozygous disruption in a gene encoding TRP -
XX Disclosure; Fig 15; 106pp; English.
XX The present sequence represents the DNA sequence of an expanded T243
XX gene. The T243 gene encodes for a trinucleotide repeat protein (TRP).
XX The invention describes methods of producing embryonic stem (ES)
XX cells comprising a heterozygous disruption in a target DNA sequence
XX (preferably gene T243) encoding a TRP and of producing a knockout mouse
XX comprising a homozygous disruption in a gene encoding TRP, where the
XX disruption inhibits the production of the wild type TRP. The invention
XX also relates to identifying agents capable of affecting a phenotype of
XX a knockout mouse. Also described are methods of determining whether
XX expansion of the trinucleotide repeat in a gene encoding TRP produces a
XX phenotypic change. The transgenic animals and the cells are useful for
XX identifying compounds capable of ameliorating disease symptoms, and as
XX test substrates for the identification of drugs, pharmaceuticals,
XX therapies and interventions which may be effective in treating

CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
CC disease. The animal models for trinucleotide repeat disorders are ideal
CC model systems to study the progression of disease in vivo, the molecular
CC basis of these diseases and show the features observed in human disease.
CC Using the mice, it is possible to model both the pathogenic mechanism and
CC the trinucleotide repeat instability in the mouse.
XX
SQ Sequence 1848 BP; 415 A; 519 C; 543 G; 371 T; 0 other;
Query Match 93.8%; Score 1725.2; DB 22; Length 1848;
Best Local Similarity 96.6%; Pred. No. 0;
Matches 1786; Conservative 0; Mismatches 53; Indels 9; Gaps 2;
Qy 1 ggcacgaggaggagagcgccgggtccgctctgctgctcgggtccggtcgccatgag 60
Db 1 ggcacgaggaggagagcgccgggtccgctctgctgctcgggtccggtcgccatgag 60
Qy 61 tccatgtctgagctcgccgcccgcgctctcttatttctctgctgctgctcttcgctg 120
Db 61 tccatgtctgagctcgctgctgctgctgctgctgctgctgctgctgctgctg 120
Qy 121 ctgctcct-----tctgccccgaagctagcccgccgagtcgccgcccgggtcaggaga 172
Db 121 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 180
Qy 173 ccg-actgggtgcgattgccagcaaatgcgaagtgtgcgaagtattgtgctgctgagagctg 231
Db 181 ctgctgctgctgctgattgccagcaaatgcgaagtgtgcgaagtattgtgctgctgagagctg 240
Qy 232 aagtcggccttttgaggaaacgggaaacacgaaggaagtgattgacacggcgtctatggcctc 291
Db 241 aagtcggccttttgaggaaacgggaaacacgaaggaagtgattgacacggcgtctatggcctc 300
Qy 292 ctgacgagggaaggctctgagtgcaagtacacccaagtcggaacttcggttattatgaagtc 351
Db 301 ctggacgggaaggctctgagtgcaagtacacccaagtcggaacttcggttattatgaagtc 360
Qy 352 actgagaccatttgcaagaggctctggaactacacgctgcacaaaggagaggaactggaagc 411
Db 361 actgagaccatttgcaagaggctctggaactacacgctgcacaaaggagaggaactggaagc 420
Qy 412 aaccggtttgccaaagggtatgtcggagagaccttgcgacgctgcacaaacctagtcacacaa 471
Db 421 aaccggtttgccaaagggtatgtcggagaccttgcgacgctgcacaaacctagtcacacaa 480
Qy 472 ggggtcaagggtggtgtaggatatccctctatgctgctggaacgagacctcagcagaggtg 531
Db 481 ggggtcaagggtggtgtaggatatccctctatgctgctggaacgagacctcagcagaggtg 540
Qy 532 gctgacctcaagaagcagctgacgtgctggtggaagagtttgaaagaggtgattgagagac 591
Db 541 gctgacctcaagaagcagctgacgtgctggtggaagagtttgaaagaggtgattgagagac 600
Qy 592 tggctacaggaaccacacagaggaagacctgactgaattccctctgctgccaacacagctgctg 651
Db 601 tggctacaggaaccacacagaggaagacctgactgaattccctctgctgccaacacagctgctg 660
Qy 652 aagggaaaggacagagttgcttagcagagcgtggtctggaagagggggacatagcc 711
Db 661 aagggaaaggacagagttgcttagcagagcgtggtctggaagagggggacatagcc 720
Qy 712 tccctggagggaaagaaatccaaagaagcgagcgaggtcctcctccatgagtcg 771
Db 721 tccctggagggaaagaaatccaaagaagcgagcgaggtcctcctccatgagtcg 780
Qy 772 agcagcaagcagaggaagaaactgggggctcctgggggaggtgcacacgcgcagagagagag 831
Db 781 agcagcaagcagaggaagaaactgggggctcctgggggaggtgcacacgcgcagagagagag 840
Qy 832 gaggggtgcgaagagcctcgcctccacacacagccccctctgtagctgtgagcccgag 891
Db 841 gaggggtgcgaagagcctcgcctccacacacagccccctctgtagctgtgagcccgag 900

Qy 892 cttagtctcttgaatcaagaccctgacttcagagcttgagacacgacagcgagcgc 951
 Db 901 cttagtctcttgaatcaagaccctgacttcagagcttgagacacgacagcgagcgc 960
 Qy 952 agcgagctcagcaagagcagctgctcagacatcaggtctcctctggtctgccc 1011
 Db 961 agcgagctcagcaagagcagctgctcagacatcaggtctcctctggtctgccc 1020
 Qy 1012 cttctctctcttgaaacacagcaagagtggaagagctcgtgggagcagcga 1071
 Db 1021 cttctctctcttgaaacacagcaagagtggaagagctcgtgggagcagcga 1080
 Qy 1072 ccccaaaaggag 1131
 Db 1081 ccccaaaaggag 1140
 Qy 1132 gagctccgggttccaccagatcccccagagctgagatccagagctcctgagagctg 1191
 Db 1141 gagctccgggttccaccagatcccccagagctgagatccagagctcctgagagctg 1200
 Qy 1192 aagagcatgagggcacataagatgctcaccagcgcccccttcagccaggaagagctccgtg 1251
 Db 1201 aagagcatgagggcacataagatgctcaccagcgcccccttcagccaggaagagctccgtg 1260
 Qy 1252 cagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1311
 Db 1261 cagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320
 Qy 1312 gtcggatgtaaatccgaaactccagacaaactcgggtgtggcacaaggggctgga 1371
 Db 1321 gtcggatgtaaatccgaaactccagacaaactcgggtgtggcacaaggggctgga 1380
 Qy 1372 cgcagagccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1431
 Db 1381 cgcagagccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1440
 Qy 1432 aagccaatctcagttcccggtgtccaccactcctcctcctcagagagcagcagcagcagcagcagc 1491
 Db 1441 aagccaatctcagttcccggtgtccaccactcctcctcctcagagagcagcagcagcagcagcagc 1500
 Qy 1492 agccctctccagggctaccagagtaaacacaccttttgcccttgcgtttggttgcctggg 1551
 Db 1501 agccctctccagggctaccagagtaaacacaccttttgcccttgcgtttggttgcctggg 1560
 Qy 1552 tctctatcagctcagagtgctccctcctcagatctttttgctttgtcccccataccc 1611
 Db 1561 tctctatcagctcagagtgctccctcctcagatctttttgctttgtcccccataccc 1620
 Qy 1612 aggggctggaaggccatcaccatcattgagaggttaacctgacgttactaggaggtgct 1671
 Db 1621 aggggctggaaggccatcaccatcattgagaggttaacctgacgttactaggaggtgct 1680
 Qy 1672 gggagcgccgggtgtgttgggttaactcactcagctggtctcagccttctaactgctg 1731
 Db 1681 gggagcgccgggtgtgttgggttaactcactcagctggtctcagccttctaactgctg 1740
 Qy 1732 agcccttatacagttcctctctgtgtgtggtgactccacgccccccacacacacacata 1791
 Db 1741 agcccttatacagttcctctctgtgtgtggtgactccacgccccccacacacacacata 1800
 Qy 1792 aaattattcagctgtttcactaactgtataaaaaaataaaaaa 1839
 Db 1801 aaattattcagctgtttcactaactgtataaaaaaataaaaaa 1848

RESULT 3

AAS21311

ID AAS21311 standard; cDNA; 1675 BP.

XX AAS21311;

XX AAS21311;

DT 24-OCT-2001 (first entry)

XX DE Human cDNA sequence encoding for PRO4409 polypeptide.
 XX KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 XX adipocyte; A-peptide; factor VIIA; gene therapy; ss.
 OS Homo sapiens.
 XX WO200140466-A2.
 PN 07-JUN-2001.
 PD 01-DEC-2000; 2000WO-US32678.
 XX 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 09-DEC-1999; 99US-0170262.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 30-DEC-1999; 99WO-US31243.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.
 XX (GETH) GENENTECH INC.
 XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WL, Zhang Z;
 XX WPI; 2001-408281/43.
 DR P-PSDB; AAU12239.
 XX
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX Claim 3; Fig 135; 813pp; English.
 XX AAS21244-AAS21518 encode for novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood

monocytes (PBMCs), or the proliferation of endothelial cells. Some of the PRO polypeptides may modulate glucose or free fatty acid uptake by skeletal muscle cells or adipocytes; or inhibit binding of A-peptide to factor VIIa. The PRO polypeptides can be used in assays to identify molecules involved in binding interactions. The polynucleotides encoding PRO polypeptides can be used to generate probes, antisense RNA/DNA, transgenic or knock out animals and can be used in gene therapy.

Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 other;

Query Match 37.6%; Score 691.6; DB 22; Length 1675;
Best Local Similarity 84.5%; Pred. No. 4.5e-169;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

[illegible]

D	b	1127	tccagaggcatccccctctcacacagccccctgatgagtctgagcccaccgcagcat	1186
Q	y	899	tccttgaatcaagagccccctgacttcagagcttgga	934
D	b	1187	cctctgtcctgagaccctgatttgaagctgagga	1222
RESULT 4				
AAS05279				
A	I	D	AAS05279 standard; cDNA; 1362 BP.	
X	C			
X	A			
X	S			
X	T			
D	T	07-SEP-2001	(first entry)	
X	X			
D	E			
X	X			
K	W			
K	W			
K	W			
X	X			
O	S			

RESULT 4

AAS05279
ID AAS05279 standard; cDNA; 1362 BP.

XX
AC AAS05279;XX
DT 07-SEP-20[illegible]

XX. human cinnarizoline repeat protein (TRP) cDNA sequence.

KW trinucleotide repeat protein, TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; human; ss

XX
OS Homo sapiens.

XX	FH	Key	Location/Quarantine

FT	CDS	
7.843		

```
ET /"lag=" d
ET /product="TRP"
```

XX PN WO200130798-A1.

XX PD 03-MAY-2001

XX
E
C

XX
07 1000 2000, Z086750-OM0002XX
PR 26-001-1999; 9905-0

PA (DELTA-) DELTAGEN INC.
XX

PI Klein R, Matthews W, Moore M, Allen KD;
yy

WPI; 2001-300473/31.

DK
XX
P-FSDB; AAU02499.

PT Novel transgenic animals useful as animal model for characterization of function of a gene encoding trinucleotide repeat proteins (TRPs),
PT contains heterozygous disruption in a gene encoding TRP -
PT

XX
PS Disclosure: Fig 11: 106pp: English

The present sequence encodes for human trinucleotide repeat protein (TRP). The invention describes methods of producing embryonic stem (ES) cells comprising a heterozygous disruption in a target DNA sequence (preferably gene W243) encoding a TRP and of producing a knockout mouse comprising a homozygous disruption in a gene encoding TRP, where the disruption inhibits the production of the wild type TRP. The invention also relates to identifying agents capable of affecting a phenotype of a knockout mouse. Also described are methods of determining whether expansion of the trinucleotide repeat in a gene encoding TRP produces a phenotypic change. The transgenic animals and the cells are useful for identifying compounds capable of ameliorating disease symptoms, and as test substrates for the identification of drugs, pharmaceuticals, therapies and interventions which may be effective in treating trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's disease. The animal models for trinucleotide repeat disorders are ideal model systems to study the progression of disease *in vivo*, the molecular basis of these diseases and show the features observed in human disease using the mice, it is possible to model both the pathogenic mechanism at the trinucleotide repeat instability in the mouse.

Sequence 1362 BP; 352 A; 376 C; 392 G; 242 T; 0 other; 55Q

Query Match

36.28; Score 665.8; DB 22; Length 1362;

[illegible]

CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 86
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX04311 for described uses).

Sequence 1657 BP; 361 A; 490 C; 503 G; 297 T; 6 other;

Query Match 24.3%; Score 446; DB 20; Length 1657;
Best Local Similarity 81.1%; Pred. NO. 2.1e-105;
Matches 542; Conservative 2; Mismatches 117; Indels 7

[illegible]

RESULT 6

AAS25051

ID AAS25051 standard; cDNA; 580 BP.

XX

AC AAS25051;

XX
XX

DT	07-NOV-2001 (first entry)
XX	
DE	Human ovarian PCR-subtracted cDNA library clone #1232.
XX	
KW	Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss; gene therapy; cystostatic; T-cell expansion; nucleic acid hybridisation;

KW primer; probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157207-A2
 XX
 PD 09-AUG-2001.

XX
PF 05-FEB-2001; 2001WO-US03733.

XX
XX

PR 04-FEB-2000; 2000US-0180403.

PR 28-MAR-2000; 2000US-

XX

PA (CORI-) CORIXA CORP.
XX

XX Algate PA.
PT Algate PA. Manton J:

XX
XX
XX
XX
PT
PT
PT

XX

Example 1; page 301-302; 378pp; English.

Sequence 580 BP: 155 A: 136 C: 172 G: 108 T: 9 other;

Query Match 22.4%; Score 412.2; DB 22; Length 580;
Best Local Similarity 90.0%; Pred. NO. 7.6e-97;
Matches 441; Conservative 0; Mismatches 49; Indels 0;

Ov 234 atcggccttttgaagaaacagggaagaccgaagaaagtgaattgacacccgactatggcatcct 293

27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054

Db 1 gtcagccttgaggaaaccggcaagaccaaggagggtgattgacacgggctatggcatcct 60

Qy 294 ggacgggaagggtctggagtcaggtaacaccaagtcggacttacggttaattgaagtcac 353

61
62

027C 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099

Qy 354 tgaaccatttgcagaagaccttctgaactacagcctgcacacaaaggaaggaactgaacacaa 413

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 121 tgagaccattgtagaggctcctggattatagcctgcacaaggagagaccggcagcaa 180

Qy 414 ccggttgccaagggtatgtcggagaccttgagacgctgcacaacctagtccacaagg 473

181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

DU LGALLLYCCAAHYGCAHYLCAHYAHCCCLLYAHACALACACAAHYG Z40

Db 361 gtacaggaaccaccaggaggagacctgactgaattctctctgcgccaccacgtgtctgaa 420
Qy 654 ggaagaagacacagattgctctagcagagcggtggtctgcaagaaggaggacatagcctc 713
Db 421 gggaaaacacacagctgctgcgcagagcancgtgccgcaagaaggaggacacagacct 480
Qy 714 cctgggagg 723
Db 481 gcccgggcgg 490

RESULT 7
AAH83696
ID AAH83696 standard; cDNA; 580 BP.
XX
AC AAH83696;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:1320.
XX
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;
KW immunogenic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200151513-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01575.
XX
PR 14-JAN-2000; 2000US-0176722.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA;
XX
DR WPI; 2001-425866/45.
XX
PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to
PT treat and diagnose cancers, particularly ovarian cancer -
XX
PS Claim 5; Page 304; 338pp: English.
XX
CC AAH82377 to AAH83878 represent human ovarian tumour-associated
CC polynucleotide sequences which encode ovarian tumour proteins. The
CC ovarian tumour protein and polynucleotide sequences have cytostatic
CC activity, and can be used in gene therapy and vaccine production. The
CC ovarian tumour proteins and polynucleotides can be used to inhibit
CC the development of cancer, particularly ovarian cancer. They can also
CC be used to diagnose the onset and progression of cancer.
XX
SQ Sequence 580 BP; 155 A; 136 C; 172 G; 108 T; 9 other;

Query Match 22.4%; Score 412.2; DB 22; Length 580;
Best Local Similarity 90.0%; Pred. No. 7.6e-97;
Matches 441; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 234 gtcggcttttgaggaaacgggaaagaccacgaagtgattgacacggcgtatgcatcct 293
Db 1 gtcagcctttgaggaaacggcgaagaccacgaagtgattgacacggcgtatgcatcct 60
Qy 294 ggaagggaaggctctggagtcagtcacacgaagtcggacttacgttaattgaagtcaac 353
Db 61 ggaccagaaggcctctggagtcacacgaagtcggacttacgttaattgaagtcaac 120
Qy 354 tgagaccatttgcaagagctcttgactacagcctgcacaaaggagagactggcgacaa 413
Db 121 tgagaccatttgcaagagctcttgattatagcctgcacaaaggagagacggcgacaa 180
Qy 414 ccggtttgcaagggtatgctcgagacaccttgagacgctgcacaaacctagtcacaaagg 473

Db 181 tgcatttccaaggcgcatgtcagagaccttgagacattacacaacctggtacacaaagg 240
Qy 474 ggtcaagtggtgatgatatccctctatgagctgtggaacgagacctcagcagaggtggc 533
Db 241 ggtcaagtggtgatgagacatccctctatgagctgtggaacgagacctctgcagaggtggc 300
Qy 534 tgacctcaagaagcagtgtagctgtggaagagtttgaaaggtgattgagagactg 593
Db 301 tgacctcaagaagcagtgtagctgtggaagagtttgaaaggtgattgagagactg 360
Qy 594 gtacaggaaccaccacaggaggagacctgactgaattctctgtgccaaaccacgtgctgaa 653
Db 361 gtacaggaaccaccacaggaggagacctgactgaattctctgtgccaaaccacgtgctgaa 420
Qy 654 gggaaaagacacagctgctctagcagagcggtggtctgcaagaaggaggacatagcctc 713
Db 421 gggaaaagacacagctgctctgagcagagcancgtgccgcaagaaggaggacacagacct 480
Qy 714 cctgggagg 723
Db 481 gcccgggcgg 490

RESULT 8
AAA42779
ID AAA42779 standard; cDNA; 274 BP.
XX
AC AAA42779;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:1519.
XX
KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
KW antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021990-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24205.
XX
PR 15-OCT-1998; 98US-0104435.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
XX
WPI; 2000-317937/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 479; 618pp; English.
XX
CC AAA4261 to AAA43419 represent specifically claimed secreted expressed

Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1	
Qy	162 ggctgagagaccgactgggtgcgattgccagcaaatgcgaagtgtgcaagtattgtgc 221

XX	(ZYMO) ZYMOGENETICS INC.		
FA	Moore EE, Raymond FC, Sheppard PO;		
PI	WPI; 1999-418753/35.		
XX	P-PSDB: AAY25421.		
DR	New mammalian nucleic acid encoding a calcitonin-like polypeptide-1		
XX	Disclosure; Page 50-52; 55pp; English.		
XX	This invention describes a novel human polynucleotide that encodes a		
CC	calcitonin-like polypeptide which has osteopathic, antidiabetic and		
CC	hypotensive activity and also acts as a vasodilator. Antibodies derived		
CC	from the product of the invention can be used to isolate and purify		
CC	the polypeptide sequences. The polypeptides can be used as		
CC	neuromodulators in a variety of peripheral organs. They can be used in		
CC	the treatment of Paget's disease, hypercalcaemia, and osteoporosis. They		
CC	can also be used to treat Raynaud's disease, and may be able to inhibit		
CC	the progression of type I diabetes. The polypeptide sequences may be		
CC	administered as vasodilators to treat hypertension.		
XX	Sequence 1457 BP; 372 A; 323 C; 440 G; 322 T; 0 other;		
SQ			
Query Match 8.7%; Score 160.2; DB 20; Length 1457;			
Best Local Similarity 57.3%; Pred. No. 2.4e-31;			
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;			
Qy	162	ggctgagagaccgactgggtgcgattcccccagcaaatgcgaagtgcgaagtattgtgc	221
Db	137	ggagagagacgatgacacagacgcttcccccagcaaatgcgaagtgcgaagtattgtgc	196
Qy	222	tgtgagctgaagtcggtcttttgcgaagacgggaaagcaccaggaagtgattgacaccgg	281
Db	197	cacagagctacagcggaactgagtcgcacccggtcgatctcgagaggtgctggagctggg	256
Qy	282	ctatggcatcctggcgggaagggctctggaagtcgaag---tacaccaagtgcgacttacg	338
Db	257	gcaggtgctgatacaggaaggaagagacacgtgcttaccagcgtttccagagacaag	316
Qy	339	gttaattgaagtcactgagaccatttgcgaagaggtcttgcactacacccctgcacaagga	398
Db	317	gctggaagagccttagagaatttatgtgacggatcctggactatagttccacgtga	376
Qy	399	gaggactggcagcaaccggtttgcgaaggggtatgtcggagacccttggagcgtgcacaa	458
Db	377	gcgcaagggctcactgagatatgcgaaggtcagagtcagaccatggcaactgaaagg	436
Qy	459	cctagtcacaaaaggggtcaaggtggtgatggatatacccttatgagctgtggaacgagac	518
Db	437	cctagtcagaaaggggtgaaggtggtatctgggacccctctggagccttgggatgacc	496
Qy	519	ctcagcagaggtggctgacctcaagaagcagtgtagctgctggtggaagatttgaaga	578
Db	497	cagcgtgaggtgcatacctccaaagacagtgtagaccattgttggaggaggttgaaga	556
Qy	579	ggtgattgagactggtgcaggaaccacccaggaagacccctgactgactgaattcctctgtgc	638
Db	557	cattgtggagactggttacttccaccatcaggagcagccctcacaaaatttctctgtga	616
Qy	639	caaccacgtgctgaagggaaggaacacagttgctcctagcagagcgttggctggcaagaa	698
Db	617	aggtcatgtgtccccagctgtgaaactgtctacaggaacttggacttggaaagga	676
Qy	699	g	699
Db	677	g	677
RESULT 14			
AAAL6621			

ID AAL16621 standard; cDNA; 1462 BP.
 AC AAL16621;
 XX 16-JUN-2000 (first entry)
 DE Human secreted protein clone pj193_5 nucleotide sequence SEQ ID NO:7.
 XX Human; secreted protein; immunostimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; ss.
 XX Homo sapiens.
 OS W0200009552-A1.
 PN 24-FEB-2000.
 XX 13-AUG-1999; 99WO-US18298.
 XX 14-AUG-1998; 98US-0096622.
 PR 17-AUG-1998; 98US-0096815.
 PR 04-SEP-1998; 98US-0099229.
 PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.
 XX (GEMY) GENETICS INST INC.
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Tracy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fechtel K;
 XX MPI; 2000-205979/18.
 DR p-PSDB; AAY94901.
 XX New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity -
 XX Claim 16; Page 470-471; 641pp; English.
 PS AAL16618 to AAL16697 encode the human secreted proteins given in
 CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin infpndent

CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAL16698 to AAL16774 represent
 CC probes for the human secreted proteins from the present invention.
 XX Sequence 1462 BP; 384 A; 316 C; 431 G; 331 T; 0 other;
 SQ
 Query Match 8.7%; Score 160.2; DB 21; Length 1462;
 Best Local Similarity 57.3%; Pred. No. 2.4e-31;
 Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
 QY 162 ggctgagagacgactgggtgctgctccacgaacaaatgcgaagtgcgaagtgtgtgc 221
 Db 160 ggaggaggagcagatgacacagaacgcttgcacgaacaaatgcgaagtgcgaagtgtgtgc 219
 QY 222 tgtgtgctgaagtcggctgttgggaaacgggaaacacgaaggaagctgattgacacccg 281
 Db 220 cacagagctacaggcggaactgagtcgcaccggtcgatctcgagaggtgctggagctggg 279
 QY 282 ctatggcatcctggagcgggaaggctctggagtcgaag---tacaccaagtcggacttacg 338
 Db 280 gcagggtgctggatagcagggaaggaagacacgctgcttacagcgtttcagagacaag 339
 QY 339 gtttaattgaagtcactgagacccatttgcgaagaggtcttgcgactacacgctgcacaagga 398
 Db 340 gctggaagaggccttagaagaatttatgtgagcggatcctggactatagtgttcaacgctga 399
 QY 399 gaggactggcagcaaccgggtttgccaaagggtatgttcggagacacctttgagacgtgcacaa 458
 Db 400 gcgcaagggtcactgagatgacgaagggtcagagtcagacacatggaacacactgaaagg 459
 QY 459 cctagtccacaaagggttcagggtggtgagtgatataccctctatgagctgtggaacagagac 518
 Db 460 cctagtgcagaagggtggaaggtggaatctgggagctccctctggaagcttgggatgagcc 519
 QY 519 ctacgacagaggtggtgacacctcaagaagcagtgtagctgctggtggaagagtttgaaga 578
 Db 520 cagcgtggaagtcacataccctcaagaagcagtgtagcagacctgttgaggaggtttgaaga 579
 QY 579 ggtgattgagagctgttcacagggaaccacaggaggagagacctgactgaattcctctgtgc 638
 Db 580 cattgtgggagactggtacttccaccatcaggagcagccctacacaaatttctctgtga 639
 QY 639 caaccacgtgctgaagggaagacacagagttgctcctagcagagcgggtggtctggaagaa 698
 Db 640 aggtcattgtctccagctgctgaaactgcatgtctacaggaaacttggactggaaagga 699
 QY 699 g 699
 Db 700 g 700
 RESULT 15
 AAY96349
 ID AAA96349 standard; cDNA; 1470 BP.
 XX
 AC AAA96349;
 XX 08-FEB-2001 (first entry)
 DE cDNA encoding a novel polypeptide designated PRO4354.
 XX Secreted protein; transmembrane protein; PRO1484; PRO4334; PRO1122;
 KW PRO1889; PRO1890; PRO1887; PRO1785; PRO4353; PRO4357; PRO4405; PRO4356;
 KW PRO4352; PRO4360; PRO4354; PRO4408; PRO5737; PRO4425; PRO5990; PRO6030;
 KW PRO4424; PRO4432; PRO4430; PRO4499; tumour; obesity; diabetes;
 KW insulinemia; kidney disorder; Bergers disease; nephropathy;
 KW Schonelein-Henoch purpura; celiac disease; dermatitis herpetiformis;
 KW Crohns disease; ss.
 XX Homo sapiens.
 OS

```

FH Key Location/Qualifiers
FT CDS 108..854
FT /*tag= a
FT /transl_except= (pos: 423..425, aa: Lys)
FT sig_peptide 108..170
FT /*tag= b
PN WO200056889-A2.
XX 28-SEP-2000.
XX 01-MAR-2000; 200WO-US05601.
XX 23-MAR-1999; 99US-0125774.
XX 23-MAR-1999; 99US-0125778.
XX 24-MAR-1999; 99US-0125826.
XX 31-MAR-1999; 99US-0127035.
XX 05-APR-1999; 99US-0127706.
XX 21-APR-1999; 99US-0130359.
XX 27-APR-1999; 99US-0131270.
XX 27-APR-1999; 99US-0131272.
XX 04-MAY-1999; 99US-0132371.
XX 04-MAY-1999; 99US-0132379.
XX 25-MAY-1999; 99US-0135750.
XX 08-JUN-1999; 99US-0138166.
XX 20-JUL-1999; 99US-0144791.
XX 03-AUG-1999; 99US-0146790.
XX 09-DEC-1999; 99US-0170262.
XX (GETH ) GENENTECH INC.
XX Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;
XX Stewart TA, Watanabe CK, Wood WI, Zhang Z;
XX WPI; 2000-628263/60.
XX P-PSDB; AAB18922.
XX Novel secreted and transmembrane polypeptides useful for diagnosing
XX tumour in a mammal, for identifying agonists and antagonists of the
XX polypeptide and for therapeutic use
XX Claim 2; Fig 27; 222pp; English.
XX The present sequence encodes a secreted or transmembrane polypeptide.
XX The specification describes polypeptides designated PRO1484, PRO4334,
XX PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
XX PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
XX PRO6030, PRO4424, PRO4422, PRO4430 and PRO499. PRO1889 polypeptide is
XX useful for diagnosing tumour in a mammal. The polypeptides, their
XX agonists and antagonists are useful treating a condition associated with
XX expression or activity of the polypeptide. Conditions treated include
XX obesity, diabetes or hyper- or hypo-insulinemia. The polypeptides are
XX capable of inducing proliferation of mammalian kidney mesangial cells
XX and are therefore useful for treating kidney disorders associated with
XX decreased mesangial cell function such as Berger's disease or other
XX nephropathies associated with Schönlein-Henoch purpura, celiac disease,
XX dermatitis herpetiformis or Crohn's disease. The nucleic acids may be used
XX to generate transgenic animals for use in development and screening of
XX therapeutically useful reagents and also for chromosome identification
XX and tissue typing.
XX Sequence 1470 BP; 373 A; 322 C; 438 G; 337 T; 0 other;

```

```

Query Match 8.7%; Score 160.2; DB 21; Length 1470;
Best Local Similarity 57.3%; Pred. No. 2.4e-31;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
QY 162 ggctgagagaccgactgggtgcgattgccagcaaatgcgaagtgtgcaagtgtgtgc 221
DB 182 ggagaggacgacgatcacagaaacgttcccccagcaaatgcgaagtgtgtaagctgtgag 241

```

```

QY 222 tgtgagctgaagtcggctttttgaggaacacgggaaagacccaaggaagtgttgacaccgg 281
DB 242 cacagagctacagcggaactgagtcgcaccggtcgatctctgagaggtgctggagctggg 301
QY 282 ctatggaatctctggacggggaagggctctggagtgcaag---tacaccaagtgcgacttacg 338
DB 302 gcaggtgctgatacacaggcaagagagacacgctgcttacagcgttttcagagacaag 361
QY 339 gtttaattgaagtcactgagaccatttgcagagggcttctggactacagcctgcacaagga 398
DB 362 gctggaagagcccttagagaaatttatgtgagcggatcctggactatagtgttcacgctga 421
QY 399 gaggactggcagcaacccggtttgccaaaggttatgtcggagaccctttgagacgctgcacaa 458
DB 422 gcgcaagggctcactgagatatgccaaaggtcagagtcagaccatggcaacactgaaagg 481
QY 459 cctagtcacaaaaggggtcaaggtggtgatggatatcccttatgagctgtggaacagagac 518
DB 482 cctagtcagaaggggtgaagtggtatctggggatccctctggagctttgggatgagcc 541
QY 519 ctacagcagaggtgctgacctcaagaagcagtgtagctgctggtgggaagatttgaaga 578
DB 542 cagcgtggaggtcacatacctcaagaagcagtgtagaccatgttggagagatttgaaga 601
QY 579 ggtgattgagagctggttacagaaacccacggaggagaaacacctgactgaattctctgtgc 638
DB 602 catttgggagactggtacttccaccatcaggagcagccctcacaaaatttctctgtga 661
QY 639 caaccacgtgctgaagggaaagacacagtgctcctagcagagcgtgtgtctggcaagaa 698
DB 662 aggtcatgtgtccacgctgtgaaactgctgtacaggaactgtgactgggaagaa 721
QY 699 g 699
DB 722 g 722

```

Search completed: May 24, 2002, 00:23:41
Job time: 9038 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:12:35 ; Search time 3953.4 Seconds
(without alignments)
9734.381 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgagggaggagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

RESULT 1
LOCUS BC013549
DEFINITION Mus musculus, RIKEN cDNA 1600025D17 gene, clone MGC:19340 IMAGE:4222133, mRNA, complete cds.
ACCESSION BC013549
VERSION BC013549.1 GI:15488835
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1882)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.

ALIGNMENTS

1	1829.4	99.5	1882	10	BC013549	BC013549 Mus muscu
2	1784.2	97.0	1909	10	AF361644	AF361644 Mus muscu
3	691.6	37.6	1422	9	BC008898	BC008898 Homo sapi
4	691.6	37.6	1422	9	BC008898	BC008898 Homo sapi
5	691.6	37.6	1720	9	BC004423	BC004423 Homo sapi
6	491.2	26.7	962	9	HSU80744	U80744 Homo sapien
7	412.2	22.4	580	6	AX198865	AX198865 Sequence
8	412.2	22.4	580	6	AX209392	AX209392 Sequence
9	331.2	18.0	913	9	BC022093	BC022093 Homo sapi
10	284	15.4	612	9	AF161347	AF161347 Homo sapi
11	186	10.1	89855	2	AC074220	AC074220 Mus muscu
12	160.2	8.7	746	6	AR142811	AR142811 Sequence
13	160.2	8.7	746	6	AX136616	AX136616 Sequence
14	160.2	8.7	747	6	BD006701	BD006701 Novel pol
15	160.2	8.7	1413	9	BC019903	BC019903 Homo sapi
16	160.2	8.7	1419	6	AX136423	AX136423 Sequence
17	160.2	8.7	1512	6	AR142810	AR142810 Sequence
18	156.6	8.5	113109	9	HS475N16	AL035587 Human DNA
19	153	8.3	1286	3	AY060987	AY060987 Drosophill
20	151.4	8.2	16499	3	AC009379	AC009379 Drosophill
21	151.4	8.2	279530	3	AE003518	AE003518 Drosophill
22	144.4	7.9	76187	2	AC015219	AC015219 Drosophill
23	84.8	4.6	129	11	G64272	G64272 17 Human Ho
24	75.2	4.1	7218	6	166494	166494 Sequence 14
25	67.8	3.7	100345	2	AC095760	AC095760 Rattus no
26	67.6	3.7	501	6	AX094199	AX094199 Sequence
27	67.6	3.7	510	6	AX079702	AX079702 Sequence
28	66	3.6	125020	9	AF429315	AF429315 Homo sapi
29	60.6	3.3	30925	2	AC107160	AC107160 Rattus no
30	60.2	3.3	125020	9	AF429315	AF429315 Homo sapi
31	60.2	3.3	198280	10	AL593843	AL593843 Mouse DNA
32	60.2	3.3	214801	2	AL596383	AL596383 Mus muscu
33	59.8	3.3	83175	2	AC098621	AC098621 Rattus no
34	59.8	3.3	140018	2	AC099090	AC099090 Rattus no
35	59.4	3.2	132764	9	AC073842	AC073842 Homo sapi
36	59.4	3.2	167854	2	AC092483	AC092483 Homo sapi
37	58.6	3.2	161334	10	AC091629	AC091629 Mus muscu
38	58.2	3.2	86430	2	AC100626	AC100626 Mus muscu
39	58	3.2	139608	2	AC103455	AC103455 Rattus no
40	58	3.2	186238	2	AC084828	AC084828 Mus muscu
41	58	3.2	198994	10	AL583886	AL583886 Mouse DNA
42	58	3.2	215914	2	AL603802	AL603802 Mus muscu
43	57.4	3.1	10732	6	E32986	E32986 Gene encodi
44	56.8	3.1	92904	2	AC107493	AC107493 Rattus no
45	56.6	3.1	97348	10	AF091216	AF091216 Mus muscu


```
|||||
Db 1605 AGTCTCCCTCATGATCTTTTGGCTTTGTCCTCCCAATCCAGGGCTGGAAAGGCAT 1664
QY 1629 caccatattgagagcttaacctctacttagaggtgctgaggagcccgaggttg 1698
Db 1655 CACCATCATTTGAGAGCTTAACCTGTCTAGTTACTAGGAGGTGCTGGGAGCCCGGGTTG 1724
QY 1689 gtttggggttaatacactcactggctctcagcctcttaacactgcagcccttaatacagtt 1748
Db 1725 GTTTGGGGTAATCACTCACTGGCTCTCAGCTTCTAACAACCTGCAGCCCTTAATACAGTT 1784
QY 1749 cttctctgttgtagctccacagcccccacacacacacacataaattatttgatgctg 1808
Db 1785 CTTTCTCTTGTGTGACTCCACGCCCCACACACACACACATAAAATATTTCGATGCTG 1844
QY 1809 ttctacactgtataaaaaaaaaaaaaa 1839
Db 1845 TTTCTAATCTGTAAAAAATAAAAAAAAAA 1875

RESULT 2
LOCUS AF361644 1909 bp mRNA linear ROD 12-MAY-2001
DEFINITION Mus musculus putative retinoic acid-regulated protein mRNA,
complete cds.
ACCESSION AF361644
VERSION AF361644.1 GI:14028766
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1909)
AUTHORS Glozak,M.A., Li,Y., Reuille,R. and Rogers,M.B.
TITLE Trapping and characterization of novel retinoic acid response
elements
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1909)
AUTHORS Glozak,M.A., Li,Y., Reuille,R. and Rogers,M.B.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2001) Biology, University of South Florida, 4202
E. Fowler Avenue, Tampa, FL 33620, USA
FEATURES
source Location/Qualifiers
1..1909
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="P19; derived from matings between C3H/He
females and male animals carrying an X chromosome derived
from a feral mouse (McBurney & Rogers, Dev. Biol., 89,
503)"
/cell_type="embryonal carcinoma"
/clone_lib="24 hr retinoic acid-treated P19 cDNA (Bouilliet
et al., Dev. Biol., 170, 420)"
89..919
/codon_start=1
/product="putative retinoic acid-regulated protein"
/protein_id="AAK52494.1"
/db_xref="GI:14028767"
/translation="MESMSELPRLFPILLPLPILLPLPAPKLGSPAGAEETDWR
LPKSECKTVAVELKSAFEKTKREVITDGYLDGKSGVYKTKSDLRLIEVET
IKRLLDYSLHKRTSGNRFKGMSETFELHNLVHKGVKVMVDIPYELWNETSAEVA
DLKKQCDVLVEEFEEVIEDYRNHQBEDLTFELCANHLVKDKTSLAERWSGKGD
ASLGGKKSKKRGVKGSSSGSKQKELGGLGEDANAEEEGVQKASPLPHSPDPDEL
"
BASE COUNT 424 a 536 c 558 g 386 t 5 others
ORIGIN

Query Match 97.0%; Score 1784.2; DB 10; Length 1909;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 9 ggaggagagccgcgggtcgtctgctggtccggtggccatggagtcatgctc 68
|||||
```

```
|||||
Db 43 GGAGGAAGCCGCCGGCTCCGCTCTGCTCTGGTCCGGCTGGGCCATGAGTCCATGTC 102
QY 69 tgagctcgcccccgcgtgctcttattcttcttctgctgctgctcttccgctgctgctcct 128
Db 103 TGAGCTCGCCCGCCGCTGCTCTATTTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 162
QY 129 tcttgcctccgaagctagggcccgagtcctcccgccgggctgagagagccgagctgggtgagatt 188
Db 163 TCTTGCCTCCCAAGCTAGGCCCGAGTCCCGCGGGGCTGAGGAGACCCACTGGGTGCCATT 222
QY 189 gccagcaaatgcaagctgtgcaagtatgttctgtgtgagctgaagtcgctcttttgaggga 248
Db 223 GCCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGCTTTTGAGGA 282
QY 249 aacgggaaagacaaaggaagtattgacacgcgctatgagcatcctcgagcgggaaaggctc 308
Db 283 AACGGAAAGACCAAGGAAGTGTGACACCGCTATGGCATCTCGACGGGAAGGGCTC 342
QY 309 tggagtcgaagtacacaaagtcggaacttacggttaattgaaagtcactgagaccatttgc 368
Db 343 TGGAGTCAAGTACACCAAGTCGGACTTTACCGTTTAATTGAACTCACTGAGACCATTTGCA 402
QY 369 gaggtcttctgactacagcctgcacaaggagagactggcagcaaccggtttgccaagg 428
Db 403 GAGGCTTCTGACTACAGCTGCACAAAGAGAGACTGGCAGCAACCGGTTTGCACAAGG 462
QY 429 tatgtggagaccttggagacgctgcacaacctagtccacaagggggtcaagggtcgaat 488
Db 463 TATGTCCGAGACCTTTTGAGACGCTGCACAACCTAGTAGTACAGGAAGGGTCAAGGTGTGAT 522
QY 489 ggatccccctatgagctgtggaacgagacctcagcagaggtgctgacacctcaagaagca 548
Db 523 GGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCA 582
QY 549 gtgtgacgtgctggtggaagagtttgaagaggtgattgagactggtcacaggaaaccacca 608
Db 583 GTGTGACGTGCTGGTGGAGAGTGTGAAGAGGTGATTGAGGACTGGTACAGGAACCACTCA 642
QY 509 ggaggagaacctgactgaatactctctgtgccaacacagctgctgaaggaaagagcacag 668
Db 643 GGAGGAAGACCTGACTCAATTCCTCTGTGCCAACACACGCTGCTGAAGGGAAGGACACGAG 702
QY 669 ttgcctagcagagcgtgtgctcgaagaaggggagacatagcctccctcgaggagagaa 728
Db 703 TTGCCTAGCAGAGCGGTGGTCTGGCAGAAGGGGGAGATAGCCTCCTCGGAGGGAACAA 762
QY 729 atccaagaagaagcgcagcgaggttcaagggtcctcctcagtggcagcagcagcagagaa 788
Db 763 ATCCAAGAAGAAGCGCAGCGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAAGCAGAGGAA 822
QY 789 ggaactgggggggctgggggagagatgccaacgcgagagagagaggggtgtgcagaaggc 848
Db 823 GGAACCTGGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGCTGTGCAGAAGGC 882
QY 849 atcgccctcccaacacagcccccctgagtgagctgagccagcttagtctccttgaatc 908
Db 883 ATGCCCTCTCCACACACACCCCTCTGATGAGCTGTGAGCCACAGCTTAAGTCTCTTGATC 942
QY 909 aagaccctgacttcagagcttggaacacacacagcgagcgcagcagcagctccagcaag 968
Db 943 AAGACCCCTGACTTCACAGCTTGGGACACGCACA -CGCAGCGCAGCGCAGCTCCAGCAAG 1001
QY 969 gacagctgctgtccagcatcaggtctcctccttggctgtgcccccttctccttgc 1028
Db 1002 GACAGCTGCTGTCCAGCATCAGGTCTCTCTCTCTGGCTGTGCCCCCTTTCTCTCCCTTGA 1061
QY 1029 caacagcaagaggtggaagatctgggggtgctgggagacggcaccacaaagggaagagga 1088
Db 1062 CAACAGCAAGAGGTGGAAGGATCTGGGGTCTGGGAGACCGGCACCCCAAGGGAAGAGGA 1121
QY 1089 ggaggagcagaaggcagctctcttctacacagtcctccctcagcagctccgggtccacc 1148
|||||
```

Db	1122	GGAGGAGCAGAAGCAGCGTCTCTTTCTATACAGTCCCCCTCACAGCTCCGGGGTCCACC	1181
Qy	1149	cagcatccccagggctgagatccaggtcctgacatggaagctgaagagcatgagccacat	1208
Db	1182	CAGCATCCCAGGGCTGAGATCCAGGCTCCTTGACATGGAAGCTGAAGAGCATGAGGCACAT	1241
Qy	1209	aagatgtctaccagcgcgccttcagccaggaaggaactccgtgacgctcagcagccagg	1258
Db	1242	AAGATGTCTACACAGCGGCCCTTCAGCAGGAAGGACTCCGTGACGCTCACAGCAGCCAGG	1301
Qy	1269	cctgcctctctctccaccagcatctctctgctggtctctggtatggttaaatctg	1328
Db	1302	CTGCGCTTCTCTCCACCAAGCATTTCTTCTGCTGGTCTTCTCGGATGGTAAATCG	1361
Qy	1329	agaacttcaggagcaaaactcgggtgtgtgcaaaaaggggctggagccagagccagagcca	1388
Db	1362	AGAACTTCCAGGAGCAAACTCGGGTGTGGCAAAAGGGGTGGAGCCGACAGCCAGAGCCCA	1421
Qy	1389	cgccagagactgcagagagggcacctgacctaaaccccccttgaaagcgaactctgcagttc	1448
Db	1422	CGCCAGAGACTGCAGAGAGGGCACCTGACCTAACCCCTCTGGAAGCCCAATCTCGAGTTC	1481
Qy	1449	ccgtgtccacccactcctcctgagagcgtcctatgctctgtccccagcccttctcccagggc	1508
Db	1482	CCGTGTCCAGCCACTCCTCTCTGAGAGCGCTCATGCTCTGCCACAGCCCTTCTCCAGGCG	1541
Qy	1509	taccagagtaaaacacttttggcccttcggtttggttccgtggtccctcatcagcctccag	1568
Db	1542	TACCAGAGTAACACACTTTTGGCCCTTTCGGTTCCTGGTTCCTCAGCAGCCTCCAG	1601
Qy	1569	agtgccctccatcgatcttttttgcctttgtcccccaaatccccaggggctggaagggcat	1628
Db	1602	AGTCTCCCTCATCGATCTTTTGGCTTTGTCCCTCATCCACAGGGTGTGGAAGGCCAT	1661
Qy	1629	cacatcattggaggcttaacctgcagttaactaggagggtgctggggagcgcgcgggggttg	1688
Db	1662	CACCATCATTTAGAGGNTAACTCTGACTTACTAGAAAGGTGCTGGGAGCGCCCGGGGTG	1721
Qy	1689	gtttggggtaactcactcagtggtctcagccttctaacaactgcagcccttaatacacagtt	1748
Db	1722	GTTTGGGTAATCACTNACTTGGCTCTCAGCCTTANAACACTGCAGCCCTTAATACAGTT	1781
Qy	1749	cctctgtgtgtgactccccagccccccacacacacacataaaattatttcgatcgtg	1808
Db	1782	CCTTCTGTGTGTGACTCCACGCCCCCACACACACACACATAAAATATTTTGGATGCTG	1841
Qy	1809	ttcataactgtaa	1822
Db	1842	TTTCATTAAGTGTA	1855
RESULT	3		
LOCUS	BC008898	1422 bp	linear
DEFINITION	Homo sapiens, clone MGC:1220	IMAGE:2959532,	mRNA, complete cds.
ACCESSION	BC008898		
VERSION	BC008898.1	GI:14286203	
KEYWORDS	MGC.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1. (bases 1 to 1422)		
JOURNAL	Direct Submission		
	Submitted (29-MAY-2001) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk		
	Email: cgabs-r@mail.nih.gov		
	Tissue Procurement: ATCC		

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate; 2 Row; m Column: 12
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6002306.

FEATURES	Location/Qualifiers
source	1..1422
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="MGC:1220 IMAGE:2959532"
	/tissue_type="Colon, adenocarcinoma"
	/clone_lib="NIH_MGC_15"
	/lab_host="DH10B-R"
	/notes="Vector: pOTB7"
CDS	78..914
	/codon_start=1
	/product="Unknown (protein for MGC:1220)"
	/protein_id="AAH08998.1"
	/db_xref="GI:114286204"
	/translation="MDSMPEPASRCRLPLPLLLLLLPAPELPQSQAERNDWVR LPKCKVCRKYAVELKSAFEETGCTKEVTGYILDQKASGVYKTSDLRLIEVET DKRLLDYSLHKERTGSNFRKMGSTFFTLHLNVHKGVKVMIDIPVLWNETSAEVA DLKKDDLVVEEPFEEIOWRHOBEDLTEFLCANHLVKGDTSCLAEOWSKKGDT AALUGCKXKKSRAAKAGRRSSSRKORLEGLGDPSPPEEDEGTQKASPLTHSPDDEL"

BASE COUNT	351 a	400 c	422 g	249 t
ORIGIN				
Query Match	37.6%;	Score 691.6;	DB 9;	Length 1422;
Best Local Similarity	84.5%;	Pred. No. 4.8e-164;		
Matches 791;	Conservative	0;	Mismatches 139;	Indels 6; Gaps 1;
Qy	5	cgaggaggaagccgcgggttcgctctgcttgctggttcgctggtcgccatggagttcca	64	
Dd	28	CGCGGAGGAGGAACGCCGGTGCTCTTTAGGGTCGGGGCCGGCCCATGATTCAA	87	
Qy	65	tgtctgagtctgcgccgcgtgcctctatttccttttgtctgctgctcttcgcgtctgc	124	
Dd	88	TGCTGAGCCGGCTCCGGCTGTCTTCTGCTTCTCCTTGTCTGCTGCTGCTGCTGC	147	
Qy	125	tccttctgccccgaactagcccgcgagtcgccgcgggctgagagacgcactgggtgc	184	
Dd	148	TGCTGCCGGCCCCGGAGCTGGCGCCGAGACCAGCCGCGGCTGAGGAGAACGACTGGGGTTC	207	
Qy	185	gattgccagcaaatgcgaagtgtgcaagtatgttgctgtaggactgaagtgcgcttttg	244	
Dd	208	GCCTGCCCAAGAAATGGAAGTGTGTAATAATGTGTGTGGACTGAAGTCAACCTTTG	267	
Qy	245	aggaaccgggaagaacaaagaaagtattgatcacaccgctatggcatcctcgacgggaagg	304	
Dd	268	AGGAACCGCAACCAAGAGGTGATTGSCACGGGCTATGGCATCCTGGACCAGAAGG	327	
Qy	305	gctctgagtcaagtacaccaagtctgaccttaattgaagtcaactgagaccatttt	364	
Dd	328	CCTCTGGAGTCAAAATACACCAAGTCGGACTTTCGGGTTAATCGAAGTCACTGAGACCATTT	387	
Qy	365	qcaagagcttcttgactacagcttcacagagagagactggcagacaacgggtttgccca	424	
Dd	388	GCAAGAGCTCCTGGATTATAGCCTTGCAACAGAGAGAGCCGCCGACCAATCATTTGCCA	447	
Qy	425	agggtatgtcggagacacctttgagacgctgcacaacaccttagtcacaaaagggtcgaagtgg	484	
bh	448	AGGCATGTCAGACACCTTTGAGACATTACACACCTTGGTACACAAGAGGGCTCAAGTGG	507	

Db 1238 CCTCTCTCTGAGACCCCTGATTTTGAAGCTGAGGA 1273

```
RESULT 6
HSU0744          HSU0744          962 bp  mRNA  linear  PRI 18-DEC-1997
LOCUS            Homo sapiens CTG4a mRNA, complete cds.
DEFINITION
ACCESSION        U80744
VERSION          U80744.1  GI:2565062
KEYWORDS
SOURCE           human.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS          Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
                  Breschei,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
                  cDNAs with long CAG trinucleotide repeats from human brain
                  Hum. Genet. 100 (1), 114-122 (1997)
TITLE            2 (bases 1 to 962)
JOURNAL          Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
                  Breschei,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
MEDLINE          Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
                  Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
REFERENCE
AUTHORS          Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
                  Breschei,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
                  Direct Submission
TITLE            Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
                  Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
JOURNAL          Location/Qualifiers
FEATURES
source           1..962
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="1"
                  /tissue_type="cerebral cortex"
gene             1..962
                  /gene="CTG4a"
CDS              388..819
                  /gene="CTG4a"
                  /note="polyleucine rich"
                  /codon_start=1
                  /product="CTG4a"
                  /protein_id="AAB91442.1"
                  /db_xref="GI:2565063"
                  /translation="MDSMPFASRCLLLPLLLLLPAPLPGSOGAEENDWVR
                  LPSCVEKYVAVELKSAFETGKTKEVIGTGYGLDOKASGVYTKRSDLRIEVTET
                  ICKRLIITACFRFGPAATDLPRACORLPHRTTWTYKSRW"
repeat_region   439..462
                  /rpt_type=tandem
                  /rpt_unit=CTG
BASE COUNT      200 a 252 c 315 g 194 t 1 others
ORIGIN
Query Match     26.7%; Score 491.2; DB 9; Length 962;
Best Local Similarity 87.3%; Pred. No. 2.5e-113;
Matches 549; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

QY 2 gcacgaggaggaagcgcgcgcggctcgtctctgggtccgcgctgggcccgcctggaggt 61
Db 335 GCCCGCGGAGGAGGAACCGCCCGTCTTTAGGCTCCGGGCGCGCGGGCCATGGATT 394
QY 62 ccattctgagctcgcgcgcgcgcctcttatttccttgcctgcctgcctgcctgc 121
Db 395 CAATGCCCTGAGCCCGCGTCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 454
QY 122 tgcctctctctcccccgaagctagccgagtcgcccgccgagtcgagagaccgactggg 181
Db 455 TGCTGCTCCCGCCCGGAGCTGGCGCCGAGCCAGCCGAGCTGAGGAGACGACTGGG 514
QY 182 tgcgattgccccagcaaatgcgaagtgtgcaagtattgtctgtgagctggaagtcggctt 241
Db 515 TTTCGCTGCCCCAGCAATGCGAAGTGTGTAAATATGTTGCTGTGGAGCTGAAGTCAGCCT 574
QY 242 ttgaggaacagggaagaccgaaggagtgattgacacacggctatggcatcctcctggacggga 301
```

```
Db 575 TTGAGGAAACCGCGCAAGACCAAGAGGTGATTGGCACGGGCTATGGCATCTCTGGACCAGA 634
QY 302 agggctctggagtcgaagtacaccagtcggacttcacggttaattgaagtcactgagacca 361
Db 635 AGGCCTCTGGAGTCAAAATACACCAAGTCGGACTTGGGTTAATCGAAGTCACTGAGACCA 694
QY 362 ttgcaagagcctctggactacagcctgcacaagagagagagactggcagcaaccggtttg 421
Db 695 TTTGCAAGAGGCTCTT-GATTATAGCCTGCACAAGAGGAGGACCGCGCAGCAATCGATTGG 753
QY 422 ccaagggtatgtcggagacacctttgagacgctgcacaacacctagtcacacaagggtcaagg 481
Db 754 CCAAGGGCATGTCAGAGACCTTTTGAGACATTACACAACCTGTGTACACAAGGGTCAAGG 813
QY 482 tggatgagatatacccttatgactgtggaacgagacacctcagcagagtggtgacctca 541
Db 814 TGGTGATGACATCCCTATGAGCTGTGGAAGGAGACTTCTGCAGAGGTGGCTGACCTCA 873
QY 542 agaagcagtgacgtcgtcgtggaagagtttgaagagtgattgaggagctggtacagga 601
Db 874 AGAANCAGTGTGATGTGCTGGTGGAGAGAGTTTGAGGAGGTGATCGAGGACTGTTACAGGA 933
QY 602 accaccaggaggaacacctgactgaattc 630
Db 934 ACCACCAGGAGGAGACCTGACTGAATTC 962

RESULT 7
AX198865          AX198865          580 bp  DNA  linear  PAT 29-AUG-2001
LOCUS            Sequence 1320 from Patent WO0151513.
DEFINITION
ACCESSION        AX198865
VERSION          AX198865.1  GI:15389191
KEYWORDS
SOURCE           human.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS          Algate,P.A.
TITLE            Ovarian tumor-associated sequences
JOURNAL          Patent: WO 0151513-A 1320 19-JUL-2001;
                  CORIXA CORPORATION (US)
FEATURES
Location/Qualifiers
source           1..580
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
BASE COUNT      155 a 136 c 172 g 108 t 9 others
ORIGIN
Query Match     22.4%; Score 412.2; DB 6; Length 580;
Best Local Similarity 90.0%; Pred. No. 2.4e-93;
Matches 441; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 234 gtcgctttgaggaacgcggaaagacccaaggaagtgattgacacccgctatggcatcct 293
Db 1 GTCACCCCTTTGAGGAAACCGCGCAAGACCCAGGAGGTGATTGACACCGGCTATGGCATCTCT 60
QY 294 ggaacgggaaggcctctggagtcgaagtacaccaagtgcggacttaacggttaattgaagtcac 353
Db 61 GGACCAGAAAGGCGCTCTGGAGTCAAAATACACCAAGTCGCAGCTTGGCGTTTAATCGAAGTCAC 120
QY 354 tgagaccatttcgaagagggctctggactacagcctgcacaaggagagactggcagcaa 413
Db 121 TGAGACCAATTTGTAAAGAGGCTCTCTGGATTATAGCTTGCAAGAGGAGACCGCGCAGCAA 180
QY 414 ccggtttccaaagggtatgtcggagacaccttgagacgctgcacaacctagtcaccaaaagg 473
Db 181 TCGATTTGCCAAGGCGCATGTCAGAGACCTTTTGAGACATTACACAACCTTGTTACACAAAGG 240
QY 474 ggtcaagctggtgatggatatccccctatgactgtggaaacgagacacctcagcagaggtggc 533
```

Db 241 GGTCAAGTGGTATGGACATCCCTATGAGCTGTGGAACGAGACTTTCGACAGGTGC 300
QY 534 tgacctcaagaagcagtgtagctgctgtggaagagtttggaagtgattgagactg 593
Db 301 TGACCTCAAGAAGACAGTGATGTGCTGTGTGGAAGAGTTTGAGGAGTGATCGAGGACTG 360
QY 594 gtacaggaaccaccaggaggaagacctgactgaattctctctgtgccaaaccagtgctgaa 653
Db 361 GTACAGGAACACCAGGAGGAAGACCTGACTGAATTCCTCTCGCCAACACCTGCTGAA 420
QY 654 gggaaaggacagagtgctgctgacagagcggtggtctggcaagaaggggacatagcctc 713
Db 421 GGGAAAGACACAGTGTGCTGCGCAGAGCANTGGTCCGCAAGAAGGGAGACACAGCCT 480
QY 714 cctgggaggg 723
Db 481 GCCCGGGCGG 490
RESULT 8
LOCUS AX209392 580 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 1232 from Patent WO0157207.
ACCESSION AX209392
VERSION AX209392.1 GI:15423815
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 580)
AUTHORS Algate,P.A. and Mannion,J.
TITLE Compositions and methods for the therapy and diagnosis of ovarian cancer
JOURNAL Patent: WO 0157207-A 1232 09-AUG-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..580
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 155 a 136 c 172 g 108 t 9 others
ORIGIN
Query Match 22.4%; Score 412.2; DB 6; Length 580;
Best Local Similarity 90.0%; Pred. No. 2.4e-93;
Matches 441; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 234 gtccgcttttgaggaaacgggaaagaccaggaagtgttgacacccgctatggcatcct 293
Db 1 GTCAGCCTTTGAGGAACCGCAAGCAAGGAGGTGATTGACACGGGCTATGGCATCCT 60
QY 294 ggaagggaaggctctgagtcagtcacacccaagtcgacttcaggttaattgaagtac 353
Db 61 GGACCAAGAAGGCTCTGGAGTCAATATACCAAGTCGAGCTTTCGGGTTAATCGAAGTCAC 120
QY 354 tgagaccattgtcaagaggctcttgactacacgcctgcacaaggagagactggcagcaa 413
Db 121 TGAGACCAATTGTGACAGGCTCTCGATTATAGCCTGCAACAGGAGGACCGGCAGCAA 180
QY 414 ccggttgcaagggtatgtcggagaccttgagacgctgcacaacctagtcacaaaagg 473
Db 181 TCGATTGGCAAGGGCATGTTCAGAGACCTTTGAGACATTACAACTCGTACAAAAGG 240
QY 474 ggtcaagggtgatgatataccctatagctgtggaacgagacctcagcagagtgcc 533
Db 241 GGTCAAGGTGGTATGGACATCCCTATGAGCTGTGGAACGAGACTTTCGACAGGTGCC 300
QY 534 tgacctcaagaagcagtgtagctgctgtggaagagtttggaagtgattgagagactg 593
Db 301 TGACCTCAAGAAGACAGTGATGTGCTGTGTGGAAGAGTTTGAGGAGTGATCGAGGACTG 360
QY 594 gtacaggaaccaccaggaggaagacctgactgaattctctctgtgccaaaccagtgctgaa 653

Db 361 GTACAGGAACACCAGGAGGAAGACCTGACTGAATTCCTCTCGCCAACACCTGCTGAA 420
QY 654 gggaaaggacagagtgctgctgacagagcggtggtctggcaagaaggggacatagcctc 713
Db 421 GGGAAAGACACAGTGTGCTGCGCAGAGCANTGGTCCGCAAGAAGGGAGACACAGCCT 480
QY 714 cctgggaggg 723
Db 481 GCCCGGGCGG 490
RESULT 9
LOCUS BC022093 913 bp mRNA linear PRI 24-JAN-2002
DEFINITION Homo sapiens, clone IMAGE:3534003, mRNA, partial cds.
ACCESSION BC022093
VERSION BC022093.1 GI:18314384
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 913)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue procurement: DCTD/DTF
cdna Library preparation: Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 11 Row: h Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.
FEATURES
source
1..913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3534003"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
<1..422
/codon_start=3
/product="Unknown (protein for IMAGE:3534003)"
/protein_id="AAH22093.1"
/db_xref="GI:18314385"
/translation="HEEYMDIPYELWNETSAAVDLKKQCDLVVEEFEEVIEDWYRN HQEDLTFLCANHVILKGDTSCLAEQWSGKRGDTPAALGGKKKKKKRAKAAGRSS SSKQRKELGGLEGDPSPDEDEGIQKASPLTHSPDEL"
CDS
228 a 263 c 268 g 154 t
BASE COUNT
ORIGIN
Query Match 18.0%; Score 331.2; DB 9; Length 913;
Best Local Similarity 84.0%; Pred. No. 7.6e-73;
Matches 388; Conservative 0; Mismatches 68; Indels 6; Gaps 1;

```
QY 479 aggtggtgatgatccccctatagctgtggaacagagacctcagcagaggtggtgacc 538
Db 10 AGTGTGTGATGGACATCCCTATGAGCTGTGGACAGAGACTTCTGCACAGGTGGCTGACC 69
QY 539 tcaagaagcagtgtagctgtggtggaagagtttgaagaggtgattgagagctggtaca 598
Db 70 TCAAGAAGCAGCTGTGATGTCTGCTGGAAGAGTTTGAGGAGGTGATCGAGGACTGGTACA 129
QY 599 ggaaccaccagagggaagaaactgactgaattctctctgtgccaaccaagtctgaaggaa 658
Db 130 GGAACACCAGGAGGAGAGACTGACTGAATTCCTCTCGCCCAACACCGTCTGAAGGAA 189
QY 659 aggacacagattgcctagcagagcgggtggtctggaagaaggagacatagctccctgg 718
Db 190 AAGACACCAGTTCCTTGGCAGAGCAGTGTGTCGCGCAAGAGAGGAGACACAGCTGCCCTGG 249
QY 719 gagggagaagaatccaaagaagaagcgagcgagtgagcag-----ggctctctccagtgga 772
Db 250 GAGGGAAGAAGTCCAGAAGAAGAGCAGCAGGAGGCGCAAGGCGGAGGAGGAGTAGCA 309
QY 773 gcagcagcagagaagaagaaactgggggctgggggagagatgccaacgcgagagagagg 832
Db 310 GCAGCAACAAGAGAGGAGTGGTGGCTTGAGGAGAGACCCAGCCCGAGGAGGATG 369
QY 833 aggggtgcagaagggatcgccctccacacacagccccctgatgagctgtgagcccg 892
Db 370 AGGCATCCAGAGAGGCATCCCTCTCACACACAGCCCCCTGTGATGAGCTGTGAGCCCCACC 429
QY 893 ttagtgccttgatcaagaagccccctgacttcagagctgtggga 934
Db 430 CAGCATCCTCTGCTCTGAGACCCCTGATTTTGAAGCTGAGGA 471

RESULT 10
AF161347
LOCUS Homo sapiens HSPC084 mRNA, partial cds. linear PRI 01-FEB-2000
DEFINITION AF161347
ACCESSION AF161347.1 GI:6841107
KEYWORDS FLI_CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 612)
AUTHORS Zhang,Q.H., Ye,M., Zhou,J., Shen,Y., Wu,X.Y., Guan,Z.Q., Wang,L.,
Fan,H.Y., Mao,Y.F., Dai,M., Huang,Q.H., Chen,S.J. and Chen,Z.
TITLE Human partial CDS cloned from cd34+ stem cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 612)
AUTHORS Zhang,Q.H., Ye,M., Zhou,J., Shen,Y., Wu,X.Y., Guan,Z.Q., Wang,L.,
Fan,H.Y., Mao,Y.F., Dai,M., Huang,Q.H., Chen,S.J. and Chen,Z.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1999) Shanghai Institute of Hematology, Shanghai
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
Shanghai 200025, P. R. China
FEATURES
Location/Qualifiers
1..612
/organism="Homo sapiens"
/codon_start=3
/product="HSPC084"
/db_xref="taxon:9606"
/clone="CBDABE11"
/cell_type="cd34+ stem cell"
/tissue_type="cord blood"
<1..431
/codon_start=3
/product="HSPC084"
/db_xref="taxon:9606"
/protein_id="AAF28907.1"
/db_xref="GI:6841108"
/translation="NETSAEVADLKQCDVLEFEFVIEDWYRNHQEEDLTBEFLCAN
HLVKQDTSCLAQWKGDTALGKKSKKRAAGAGGSSSKQKRLGGGLRE
TPAPRRMRASRRHPLSHTAPLMSEPTQHPVLRPSSLKSE"
BASE COUNT 201 a 151 c 169 g 91 t
ORIGIN
```

```
Query Match 15.4%; Score 284; DB 9; Length 612;
Best Local Similarity 86.7%; Pred. No. 6.6e-61;
Matches 338; Conservative 0; Mismatches 45; Indels 7; Gaps 2;

QY 509 ggaacgagactcagcagagtggtgctgacctcaagaagcagctgtgacgtgctggtggaag 568
Db 1 GGAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGACAGCAGTGTGATGCTGCTGGTGGNAG 60
QY 569 agtttgaagaggtgattgagagctgggtacaggaaaccacccagggaggaagacctgactgaat 628
Db 61 AGTTTGAAGAGGTGATCGAGGACTGGTACAGAAACACACAGGAGGAGACCTGACTGAAT 120
QY 629 tcctctgtgccaaccacagctgctgaaggaagagacagagttgctctagcagagcggtggt 688
Db 121 TCCTCTCGGCCACACAGTGTCTGAAGGGAAGAAACACACAGTTGCTGTCAGAGGAGTGGT 180
QY 689 ctggaagaagggggagacatagctccctgggaggggaagaaatccaaagaagaagcgagcg 748
Db 181 CCGCAGAGAGGAGAGACACAGCTGCCCTGGGAGGAGAGAGTCCCAAGAGAAGACGACGA 240
QY 749 gagtcaag-----ggctctctccagtgctgagcagcagagagagaaagaaactggggggcc 802
Db 241 GGGCAAGGAGCAGCAGCGCGGAGGAGTAGCAGCAGCAAAACAAAGGAGGAGCTGGTGGCT 300
QY 803 tggggggaggtatgcaaacgcgagagagagaggggtgtgcagaagagcagctgccccccac 862
Db 301 TGAGGGA-GACCCCGAGCCCGGAGGAGGTAGGGGCATCCAGAGGCAATCCCTCTTCACAC 359
QY 863 acagccccctgatgagctgtgagcccg 892
Db 360 ACACCCCCCTGATGAGCTCTGAGCCCCACC 389

RESULT 11
AC074220 89855 bp DNA linear HTG 25-JAN-2002
LOCUS Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
DEFINITION WORKING DRAFT SEQUENCE, 38 unordered pieces.
AC074220
AC074220.7 GI:13876390
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 89855)
AUTHORS Han,J., Montgomery,K.T., Grills,G., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereta,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE High Throughput Mouse Sequencing
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 89855)
AUTHORS Han,J., Montgomery,K.T., Grills,G., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereta,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
COMMENT On Apr 28, 2001 this sequence version replaced gi:11276104.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site:
http://wchanning.bwh.harvard.edu:9088/hpcgs/jsp/hpcgs/sequence/mous
e.html
Contact: gnktm@acecod.bwh.harvard.edu
-----Summary Statistics
Center project name: ADY
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
```

*Consensus quality: 79355 at least Q20
*Consensus quality: 75365 at least Q30
*Consensus quality: 69022 at least Q40
**Estimated insert size: agarose-FP - N/A
**Estimated insert size: 89115 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7962: contig of 7962 bp in length
* 7963 7982: gap of unknown length
* 7983 13924: contig of 5942 bp in length
* 13925 13944: gap of unknown length
* 13945 18990: contig of 5046 bp in length
* 18991 19010: gap of unknown length
* 19011 22501: contig of 3491 bp in length
* 22502 22521: gap of unknown length
* 22522 27077: contig of 4556 bp in length
* 27078 27097: gap of unknown length
* 27098 28819: contig of 1722 bp in length
* 28820 28839: gap of unknown length
* 28840 31224: contig of 2385 bp in length
* 31225 31244: gap of unknown length
* 31245 35093: contig of 3849 bp in length
* 35094 35113: gap of unknown length
* 35114 38611: contig of 3498 bp in length
* 38612 38631: gap of unknown length
* 38632 42261: contig of 3630 bp in length
* 42262 42281: gap of unknown length
* 42282 44130: contig of 1849 bp in length
* 44131 44150: gap of unknown length
* 44151 47059: contig of 2909 bp in length
* 47060 47079: gap of unknown length
* 47080 49496: contig of 2417 bp in length
* 49497 49516: gap of unknown length
* 49517 52237: contig of 2720 bp in length
* 52238 52256: gap of unknown length
* 52257 54270: contig of 2014 bp in length
* 54271 54291: gap of unknown length
* 54291 56271: contig of 1981 bp in length
* 56272 56291: gap of unknown length
* 56292 58849: contig of 2558 bp in length
* 58850 58869: gap of unknown length
* 58870 60621: contig of 1752 bp in length
* 60622 60641: gap of unknown length
* 60642 62255: contig of 1614 bp in length
* 62256 62275: gap of unknown length
* 62276 63088: contig of 810 bp in length
* 63089 63106: gap of unknown length
* 63107 65440: contig of 2335 bp in length
* 65441 65460: gap of unknown length
* 65461 67629: contig of 2169 bp in length
* 67630 67649: gap of unknown length
* 67650 69178: contig of 1529 bp in length
* 69179 69198: gap of unknown length
* 69199 70095: contig of 897 bp in length
* 70096 70115: gap of unknown length
* 70116 71821: contig of 1706 bp in length
* 71822 71841: gap of unknown length
* 71842 73506: contig of 1665 bp in length
* 73507 73525: gap of unknown length
* 73526 74749: contig of 1223 bp in length
* 74750 74769: gap of unknown length
* 74770 75963: contig of 1194 bp in length
* 75964 75983: gap of unknown length
* 75984 77905: contig of 1922 bp in length

* 77906	77925: gap of unknown length	
* 77926	79236: contig of 1311 bp in length	
* 79237	79256: gap of unknown length	
* 79257	81332: contig of 2076 bp in length	
* 81333	81352: gap of unknown length	
* 81353	83169: contig of 1817 bp in length	
* 83170	83189: gap of unknown length	
* 83190	84096: contig of 907 bp in length	
* 84097	84116: gap of unknown length	
* 84117	85749: contig of 1633 bp in length	
* 85750	85769: gap of unknown length	
* 85770	86770: contig of 1001 bp in length	
* 86771	86790: gap of unknown length	
* 86791	88207: contig of 1417 bp in length	
* 88208	88227: gap of unknown length	
* 88228	88976: contig of 749 bp in length	
* 88977	88996: gap of unknown length	
* 88997	89855: contig of 859 bp in length.	
	Location/Qualifiers	
	1. .89855	
	/organism="Mus musculus"	
	/strain="C57BL6/J"	
	/db_xref="taxon:10090"	
	/chromosome="17"	
	/clone="RP23-76116"	
	/sex="male"	
	1. .7962	
	/note="assembly_name:Contig145"	
misc_feature	7983..13924	
	/note="assembly_name:Contig144"	
misc_feature	13945..18990	
	/note="assembly_name:Contig143"	
misc_feature	19011..22501	
	/note="assembly_name:Contig142"	
misc_feature	22522..27077	
	/note="assembly_name:Contig141"	
misc_feature	27098..28819	
	/note="assembly_name:Contig140"	
misc_feature	28840..31224	
	/note="assembly_name:Contig139"	
misc_feature	31245..35093	
	/note="assembly_name:Contig138"	
misc_feature	35114..38611	
	/note="assembly_name:Contig137"	
misc_feature	38632..42261	
	/note="assembly_name:Contig136"	
misc_feature	42282..44130	
	/note="assembly_name:Contig135"	
misc_feature	44151..47059	
	/note="assembly_name:Contig134"	
misc_feature	47080..49496	
	/note="assembly_name:Contig133"	
misc_feature	49517..52236	
	/note="assembly_name:Contig132"	
misc_feature	52257..54270	
	/note="assembly_name:Contig131"	
misc_feature	54291..56271	
	/note="assembly_name:Contig130"	
misc_feature	56292..58849	
	/note="assembly_name:Contig129"	
misc_feature	58870..60621	
	/note="assembly_name:Contig128"	
misc_feature	60642..62255	
	/note="assembly_name:Contig127"	
misc_feature	62276..63085	
	/note="assembly_name:Contig126"	
misc_feature	63106..65440	
	/note="assembly_name:Contig125"	
misc_feature	65461..67629	
	/note="assembly_name:Contig124"	
misc_feature	67650..69178	
	/note="assembly_name:Contig123"	
misc_feature	69199..70095	

[illegible]

DP	132 C
QY	282 C

[illegible]

QY 339 gtttaattgaagtcactgagaccatttgcgaagggcttctggactacagcctgcacaaagga 398
Db 312 GCTGGAAGAGGCGCTTAGAGAAATTTATGTGAGCGGATCCTGGACTATAGTGTTCACGCTGA 371
QY 399 gaggactgacagcaaccgggttgcgaaggtatgtcggagacctttgagacgctgcacaa 458
Db 372 GCGCAGGGCTCACTAGATATGCCAAGGGTCAGATCAGACCATGGCAACACTGAAAGG 431
QY 459 cctagtcccaaaagggtcaaggtggtgatgatatacccttatgagctggtggaacgagac 518
Db 432 CTTAGTGCAGAGAGGGGTGAAGGTGATCTGGGATCCCTCTCGAGCTTTGGATGAGCC 491
QY 519 ctcagcagaggtggtgcactcaagaagcagtgctgctggtggaagagtttgaaga 578
Db 492 CAGCGTGGAGGTCACATACCTCAAGAAGCAGTGTGAGACCATGTTGGAGAGCTTTGAAGA 551
QY 579 ggtgattgaggactggtacagaagaccaccaggaagagacctgactgaattcctctgtgc 638
Db 552 CATGTGGAGACTGGTACTTCCACATCAGGAGCAGCCCCCTACAAAATTTTCTCTGTGA 611
QY 639 caaccacgtgctgaagggaagacacaggttgcctagcagagcggtggtctggaagaa 698
Db 612 AGGTATGTGCTCCAGCTGCTGAARACTGCATGCTACAGGAARACTTGGACTGGAAGGA 671
QY 699 g 699
Db 672 G 672

RESULT 14
BD006701
LOCUS BD006701 747 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel polypeptide.
ACCESSION BD006701
VERSION BD006701.1 GI:18635072
KEYWORDS JP 2001029090-A/4.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 747)
Ito,Y., Mogi,S., Tanaka,H., Okubo,S. and Ogi,K.
Novel polypeptide
Patent: JP 2001029090-A 4 06-FEB-2001;
TAKEDA CHEMICAL INDUSTRIES LTD
OS Homo sapiens (human)
PN JP 2001029090-A/4
PD 06-FEB-2001
PF 19-MAY-2000 JP 2000147530
PR YASUAKI ITO,SHINICHI MOGI,HIDEYUKI TANAKA,SHOICHI OKUBO, PI
KAZUHIRO OGI
PC C12N15/09,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61K48/00,
PC A61P1/00,
PC A61P5/00,A61P9/00,A61P11/00,A61P25/00,A61P25/18,A61P31/00, PC
A61P35/00,
PC A61P37/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,G01N33/50,
PC G01N33/566//
PC C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/08,(C12P21/02, PC
C12R1:91).
PC (C12N5/10,C12R1:91),(C12P21/08,C12R1:91),C12N15/00,A61K37/02,
PC C12N5/00,
PC (C12N5/00,C12R1:91)
CC
FH Key Location/Qualifiers
FT Source 1. :747
FT /organism='Homo sapiens (human)'.
FEATURES
Location/Qualifiers
1. :747
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 209 a 143 c 255 g 140 t

ORIGIN

Query Match 8.7%; Score 160.2; DB 6; Length 747;
Best Local Similarity 57.3%; Pred. No. 1.4e-29;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
QY 162 ggcgtgaggagacgactgggtgctgattgccagcaaatgcgaagtgtgcaagtattgttc 221
Db 75 GGAGGAGGAGCGATGACACAGAACGCTTGCACAGAAATGCGAAGTGTGTAAGCTGCTGAG 134
QY 222 tgtgagactgaagtcgcgcttttgaggaaaacgggaaagacccaagaagtattgacaccgg 281
Db 135 CACAGAGCTACAGCGCGAACTCAGTGCACCGCTCGATCTCGAGAGGTGCTGGAGCTGGG 194
QY 282 ctatggcatccctggagcggaagggtcctggtgagtcag---tacaccaagtcggacttacg 338
Db 195 GCAGGTGCTGGATACAGGCAAGAGGAAGACACGTCGCTTACAGCGTTTCAGAGACAAG 254
QY 339 gtttaattgaagtcactgagaccatttgcgaagggcttctggactacagcctgcacaaagga 398
Db 255 GCTGGAAGAGGCTTAGAGAAATTTATGTGAGCGGATCCTGGACTATAGTGTTCACGCTGA 314
QY 399 gaggactgacagcaaccgggttgcgaaggtatgtcggagacctttgagacgctgcacaa 458
Db 315 GCGCAAGGGCTCACTAGATATGCCAAGGGTCAGATCAGACCATGGCAACACTGAAAGG 374
QY 459 cctagtcccaaaagggtcaaggtggtgatgatatacccttatgagctggtggaacgagac 518
Db 375 CTTAGTGCAGAGAGGGGTGAAGGTGATCTGGGATCCCTCTGGAGCTTTGGATGAGCC 434
QY 519 ctcagcagaggtggtgcactcaagaagcagtgctgctggtggaagagtttgaaga 578
Db 435 CAGCGTGGAGGTCACATACCTCAAGAAGCAGTGTGAGACCATGTTGGAGAGTTTGAAGA 494
QY 579 ggtgattgaggactggtacaggaaccaccaggaaggaacctgactgaattcctctgtgc 638
Db 495 CATTTGGGAGACTGGTACTTCCACCATCAGGAGCAGCCCCCTACAAAATTTTCTCTGTGA 554
QY 639 caaccacgtgctgaagggaaggaacacaggtgctcctagcagagcggtgctcggcaagaa 698
Db 555 AGGTATGTGCTCCAGCTGCTGAARACTGCATGCTACAGGAARACTTGGACTGGAAGGA 614
QY 699 g 699
Db 615 G 615

RESULT 15
BC019903
LOCUS BC019903 1413 bp mRNA linear PRI 22-JAN-2002
DEFINITION Homo sapiens, clone IMAGE:4991480, mRNA, partial cds.
ACCESSION BC019903
VERSION BC019903.1 GI:18043798
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1413)
Strausberg,R.
Direct Submission
Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapsb-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LUNL)
DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 42 Row: j Column: 23.

FEATURES

Location/Qualifiers
1..1413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4991480"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
c1..774

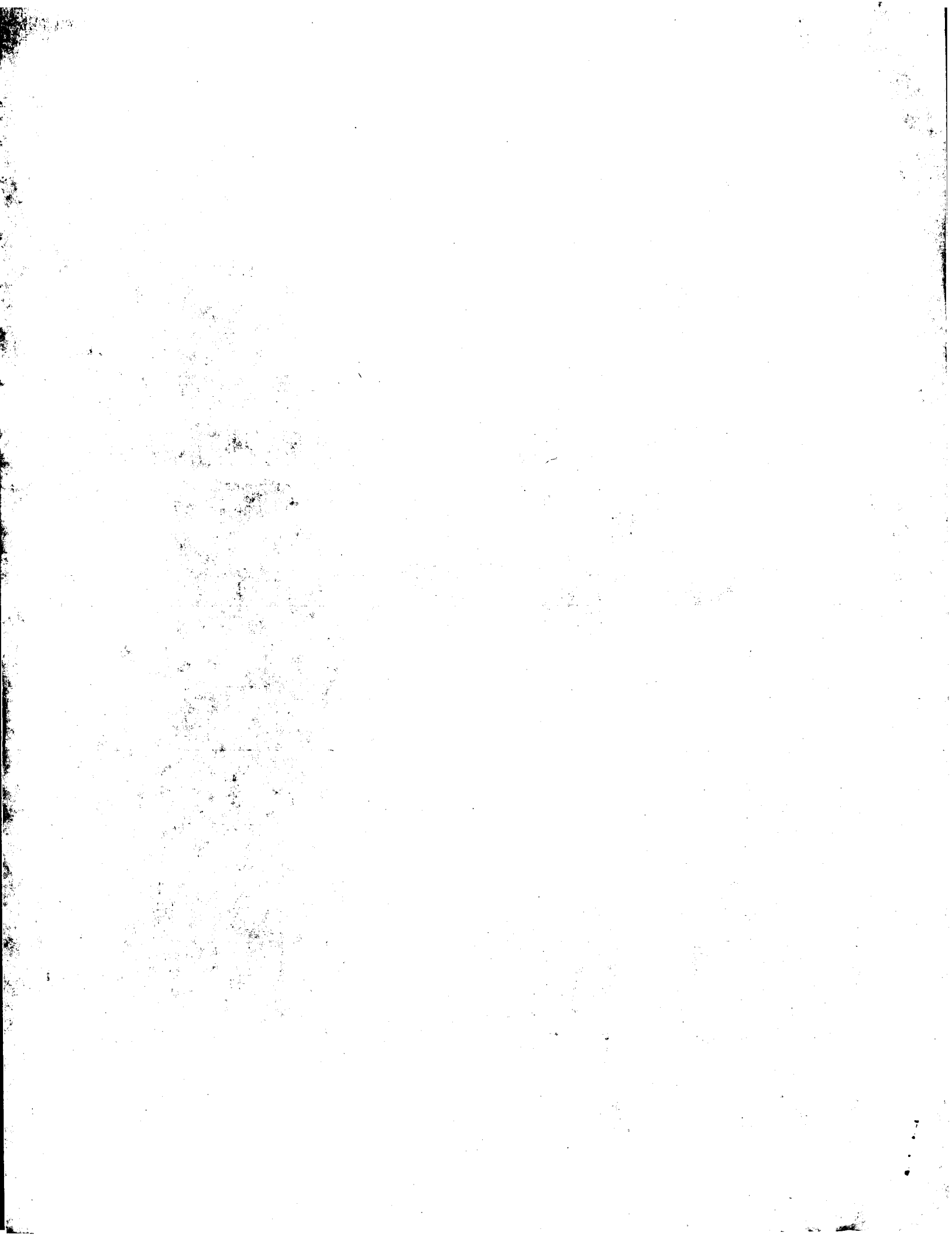
CDS

/codon_start=1
/product="Unknown (protein for IMAGE:4991480)"
/protein_id="AAH19903.1"
/db_xref="gi:18043799"
/translation="OAVIVCRRRFVMPVRLGILLFLFVAVHEAWGMLKEEDDDTER
LPSKCEVKLLSTELQAELESTGRSREVLELQVLDTGKRRHVPYSVSETRLEALE
NLCEIRLDYSVHAERKSLRYAKGQSQTMAVLKGLVQKVRKVDIGIPIELMDEPSVEV
TYLKKQCTEMLDEEDIVGDWYFHQSQPLNFCHEGVLPAETAACLOETWTGKEIT
DGEKTECEEEEEEDEEGGDKMTKTSHPKLDREDL"
BASE COUNT 378 a 301 c 413 g 321 t
ORIGIN

Query Match	8.7%	Score 160.2;	DB 9;	Length 1413;
Best Local Similarity	57.3%	Pred. No. 1.4e-29;		
Matches 310;	Conservative 0;	Mismatches 228;	Indels 3;	Gaps 1;
Qy	162	ggctgagagacccgactgggtgctgattgcccagcaaatgcaagtgatgcaagtatgtgc	221	
Db	108	CGAGGAGCAGNTGACAGAACGGTTGCCACGAATGCCGAAGTGTGTAACTGCTGAG	167	
Qy	222	tgtgagctgaagtcggcttttgaggaaacggaagaccaggaagtgattgacacccgg	281	
Db	168	CACAGAGCTACAGGCGGAACAGTGCACCGGTCGATCTCGAGAGGTGCTGGAGCTGG	227	
Qy	282	ctatggatcctggaaggggctctggagtaag---tacaccaagtcggacttacg	338	
Db	228	GCAGGTGCTGCATACAGCAAGAGGAGAGACACGTGCTTACAGCGTTTCAGAGACAAG	287	
Qy	339	gttaattgaagtcactgagaccatttgaagggctctggaactacagcctgcacaa	398	
Db	288	CTGGAAGAGCCCTAGAGAAATTTATGTGACGGGATCTGGACTATAGTTTCACGCTGA	347	
Qy	399	gaggactggcagcaaccgggtttgccaaagggtatgtcgagacacctttgagacgctgcacaa	458	
Db	348	CGCAAGGGCTCACAGATATGCCAAGGGTCAGAGTCAGACCATGCCAACACTGAAGG	407	
Qy	459	cctagtcacaaaggggtcaagggtggtgatgatatcccttatgagctgtggaacgagac	518	
Db	408	CTAGTCGAGAGGGGGTCAAGGGTGGATCTGGGGATCCCTCTGGAGCTTTGGGATGAGCC	467	
Qy	519	ctcagcagaggtggtgaccccaagaagcagtgacgtgctgtggtggaagatttgaaga	578	
Db	468	CAGCTGGAGGTCACATACCTCAAGAACGAGTGTGAGACCATGTTGGAGGATTTGAAGA	527	
Qy	579	ggtgattgagactgggtacaggaaccaccaggaaggaagacacctgactgaattcctctgtgc	638	
Db	528	CATTCGGAGACTGCTACTCCACCATCAGGAGCAGCCCTACAAAATTTCTCTGTGA	587	
Qy	639	caaccacgtgctgaaggaaggaagacacgagttgctctagacagcggtggtctgcaagaa	698	
Db	588	AGGTCATGCTGCCAGCTGCTGAAACTGCATGTCTACAGGAAACTGGACTGGAAGGA	647	

Qy 699 g 699
Db 648 G 648

Search completed: May 24, 2002, 00:14:24
Job time: 11370 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 23:05:57 ; Search time 4203.75 seconds
(without alignments)
1512.237 Million cell updates/sec

Title: US-09-696-686-50
Perfect score: 471

Sequence: 1 acagaaacagaacacaaa.....atggagtcctgctgagct 471

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	295.2	62.7	558	12	BH332134
2	159.8	33.9	416	9	AW245217
3	159.8	33.9	515	10	BE393833
4	159.8	33.9	676	10	BE383072
5	159.8	33.9	690	10	BI835105
6	159.8	33.9	828	10	BG704841
7	159.8	33.9	848	10	BE908458
8	159.8	33.9	1602	11	BC011767
9	159.6	33.9	695	10	BI760437
10	158.2	33.6	615	9	AL134371
11	158.2	33.6	695	10	BI906932
12	158.2	33.6	705	10	BI766238
13	158.2	33.6	827	10	BG827636
14	158.2	33.6	951	10	BG755172
15	156.2	33.2	724	10	BI523540
16	153.4	32.6	646	12	AG048758
17	152.8	32.4	822	10	BI770517

18	149.4	31.7	796	10	BG116444
19	147.8	31.4	799	10	BI114340
20	145.8	31.0	828	9	AL524223
21	144.8	30.7	318	12	BH332555
22	142.4	30.2	780	10	BI755826
23	141	29.9	630	9	AL534027
24	128.8	27.3	405	9	AA065263
25	122	25.9	405	9	AA065262
26	107.8	22.9	483	9	BB857917
27	107	22.7	679	10	BG762655
28	103.4	22.0	925	9	AL520942
29	103	21.9	362	10	R24856
30	100.2	21.3	898	10	BE741659
31	100	21.2	643	9	BB661603
32	98	20.8	599	10	BF780486
33	94.4	20.0	472	9	BB863087
34	94.4	20.0	776	10	BG864846
35	93	19.7	637	9	BB650071
36	93	19.7	649	9	BB652011
37	91	19.3	1553	11	AK005532
38	90	19.1	1393	10	BF316156
39	90	19.1	1586	11	AK010702
40	89	18.9	623	9	BB661026
41	87.4	18.6	502	9	BB853356
42	86	18.3	821	10	BI653380
43	85	18.0	330	9	BB856691
44	84.8	18.0	358	9	AW357718
45	83	17.6	838	10	BF534535

ALIGNMENTS

RESULT 1

BH332134
LOCUS BH332134 558 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-201K14, TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-201K14, DNA sequence.
ACCESSION BH332134 GI:17262848
VERSION BH332134.1
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 558)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartabeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment Unpublished (1999)
Other_GSSs: CH230-201K14.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 201 row: K column: 14
Seq primer: SP6
Class: BAC ends.

LOCATION/Qualifiers
1. .558
/organism="Rattus norvegicus"
/strain="BN/SSNHsd/MCW"

FEATURES
source

www-bio.lnlnl.gov/bbrp/image/image.html Base Calling / Quality
 Scores: PHRED from University of Washington Genome
 Trimming: cross_match from University of Washington
 PHRAP suite, poly_t Identification: patMatch.pl from Berkeley
 Drosophila Genome Project, University of Washington
<http://www.genome.washington.edu>
 plate: LiCM2 row: C column: 21
 High quality sequence stop: 361.

FEATURES	Location/Qualifiers
source	1. .416

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2819660"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"

```

```

/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pORG7; Site_1: xhoI; Site_2:
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into pCp1301 using the following 5'

```

cloned into EcoRI/NotI sites using the pGEM-T vector as adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, San Diego) and RNA electrophoresis kit

California, Berkeley) using ZAP-CONA systems kit (Stratagene) and Superscript II RT (Life Technologies).¹

33.9%; Score 159.8; DB 9; Length 416;
71.6%; Pred. No. 2.8e-25;

Derivative	0	Mismatches	88	Indels	7	Gaps
cgctataggcgccaaagccaccatccgctctctgattgggtgagatgg						

TTGCATAGTTCGCTGAGGCCACCATCTGCTCTTACTTGCGCAAGGCGC 148
gagatgtagtctctcattgctataagcagacgcgagcggaaccattgg 256

3'------TCCTCCCATTAGCTAGAGACAAACCC-AGAAAGCCTATTGG 201

gcggcgcccttggtcggtttcgcaagccgctagagctaccggcgagggg 316

CGCGGGCGCTTGGTCGCTTGAAGCGGGCTCGGCTCGGAGAGAGG 261
tcttcgcggttgcgctgattaccagagacagtcgcgagtcctccgaagcgg 376

TTTAGCTGTTGTGTTGCTCGGAGGACAGTGTCAGTCCCGGAAGCGG 321
tttctcgcgcgcgcgtgccgaggaagcgccgcgggtccgcctctgctct 436

27TTGCTCCGGCGCGCGGAGGAGAACCGCCGGTCCTTTAGGGTCC 381
 gggccatgggagtcctatgtctgagc 470

GGGCCATGGATTCAATGCCTGAGC 415

515 bp mRNA linear EST 21-JUL-2000
7F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3658944 5',

uence.

.1 GI:9339198

iens
a; Metazoa; Chordata; Vertebrata; Euteleostomi;
; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

S 1 to 515)
<http://mgc.nci.nih.gov/>.
 Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC348 row: n column: 01
High quality sequence stop: 513.
Location/Qualifiers
1. .515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

FEATURES source

1. .515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

78 a 158 c 181 g 98 t

Query Match

Best Local Similarity 33.9%; Score 159.8; DB 10; Length 515;
Matches 239; Conservative 0; Mismatches 88; Indels 7; Gaps 2;

QY 137 gccagtgcatacgtatagggcccaagccaccatccgtctctgattgggtgagatgg 196
Db 53 GGAAGAGACTTGGCCATAGTGCCTAGAGCCACCATCTCTTACTTGGCCAAAGGCG 112
QY 197 gaaaaaaaagatagttctctctatagcagacgagcgcgagcgaaccattgg 256
Db 113 TAAAAAGATAG-----TCCTCCCATTAGCTAGAGAGCAAAACCC-AGAAAGCCTATTGG 165
QY 257 ttgngtcgcgcgcgcgttgctggtttaccagagacacgtacccggcgagggg 316
Db 166 CTGCGCGCTCCCGGGCTTGTGCTGCTTGAAGCGGGCTCGAGAGAGGCGC 225
QY 317 cgggcggagctgcgcgcgttgctggtttaccagagacacgtacccggcgaggg 376
Db 226 GGGCGGAGGCTAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 285
QY 377 ccgggggaagctgctccgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 436
Db 286 CGAGGGGAAACTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 345
QY 437 ggggtccgctgggcccattgagtcctgagc 470
Db 346 GGGCCCG 379

RESULT 4

BE383072
LOCUS 60129888f1 NIH_MGC_19 676 bp mRNA linear EST 21-JUL-2000
DEFINITION mRNA sequence.
ACCESSION BE383072
VERSION BE383072.1 GI:9328437
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 676)

AUTHORS

NIH-MGC http://mcc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC314 row: b column: 22
High quality sequence stop: 670.
Location/Qualifiers
1. .676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library."

FEATURES

source
1. .676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library."

BASE COUNT

125 a 193 c 225 g 133 t

Query Match

Best Local Similarity 33.9%; Score 159.8; DB 10; Length 676;
Matches 239; Conservative 0; Mismatches 88; Indels 7; Gaps 2;

QY 137 gccagtgcatacgtatagggcccaagccaccatccgtctctgattgggtgagatgg 196
Db 53 GGAAGAGACTTGGCCATAGTGCCTAGAGCCACCATCTCTTACTTGGCCAAAGGCG 112
QY 197 gaaaaaaaagatagttctctctatagcagacgagcgcgagcgaaccattgg 256
Db 113 TAAAAAGATAG-----TCCTCCCATTAGCTAGAGAGCAAAACCC-AGAAAGCCTATTGG 165
QY 257 ttgngtcgcgcgcgcgttgctggtttaccagagacacgtacccggcgagggg 316
Db 166 CTGCGCGCTCCCGGGCTTGTGCTGCTTGAAGCGGGCTCGAGAGAGGCGC 225
QY 317 cgggcggagctgcgcgcgttgctggtttaccagagacacgtacccggcgaggg 376
Db 226 GGGCGGAGGCTAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 285
QY 377 ccgggggaagctgctccgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 436
Db 286 CGAGGGGAAACTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 345
QY 437 ggggtccgctgggcccattgagtcctgagc 470
Db 346 GGGCCCG 379

RESULT 5

BI835105
LOCUS 603087770f1 NIH_MGC_120 690 bp mRNA linear EST 04-OCT-2001
DEFINITION mRNA sequence.
ACCESSION BI835105
VERSION BI835105.1 GI:15946655
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/. 1 (bases 1 to 690)
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov
COMMENT	Tissue Procurement: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1570 row: 1 column: 18 High quality sequence stop: 687.
FEATURES	Location/Qualifiers 1..690 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5226785" /clone_lib="NIH_MGC_120" /lab_host="DH10B"
source	/note="Organ: pooled pancreas and spleen; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (in vitro). Research Genetics tracking code 025. Note: This is a NIH_MGC Library."
BASE COUNT	125 a 203 c 225 g 137 t
ORIGIN	
Query Match	33.9%; Score 159.8; DB 10; Length 690;
Best Local Similarity	71.6%; Pred. NO. 2.8e-25;
Matches 239; Conservative	0; Mismatches 88; Indels 7; Gaps 2;
QY 137	gacagtatactatgcattagggcgcaaacaccacctccgtctctgatgggtgagatgg 196
DB 78	GGAAGGACTTTGCCATAGGTGGCTGAGCGCCACCATCTCTCTTACTTGCCCAAGGGCG 137
QY 197	gaaaaaaaagatagttccttcatttgcgtataagaagacgacgacgacgccattgg 256
DB 138	TAAAAAGATAG-----TCCTCCATTAGTAGAGCAAAACCCC-AGAAAGCCTATTGG 190
QY 257	ttagtgccgcggcgccgttgtcttcgcaagccgctagaggctaccggcgagggg 316
DB 191	CTGCGCGCTCGCGGGGCTTGGTCCGCTTTGAAGCGGGCTCGCGCTCGAGAGAGGGC 250
QY 317	cggcgccgagctgcgttcggttatccagagacacgtcgcagtcaccgcgaagcgg 376
DB 251	GGCGGGAGGCTAGCTTGTCTGTGGTTGTCGAGGACACGTGTGACGTCCCGGAAGCGG 310
QY 377	ccgggggaagctgctccgcgcgcgtgcgcgaggaagaccccgggtccgctctgctc 436
DB 311	CGAGGGGAAACTGCTCCGCGCGCGCGGGAGGAGAACCCCGCCGGTCCITTAGGGTCC 370
QY 437	gggtccgctggccatggagtcattctctgagc 470
DB 371	GGGCCCCGGCGGCCCATGGATTCAATGCTTGACC 404
RESULT	6
LOCUS	BG704841 828 bp mRNA linear EST 07-MAY-2001
DEFINITION	602688484F1 NIH_MGC_95 Homo sapiens CDNA clone IMAGE:4820805 5', mRNA sequence.
ACCESSION	BG704841
VERSION	BG704841.1 GI:13978583

	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/. 1 (bases 1 to 828)
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov
COMMENT	Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10726 row: h column: 22 High quality sequence stop: 716.
FEATURES	Location/Qualifiers 1..828 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4820805" /clone_lib="NIH_MGC_95" /tissue_type="hippocampus" /lab_host="DH10B"
source	/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTFVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT	151 a 242 c 267 g 168 t
ORIGIN	
Query Match	33.9%; Score 159.8; DB 10; Length 828;
Best Local Similarity	71.6%; Pred. No. 2.9e-25;
Matches 239; Conservative	0; Mismatches 88; Indels 7; Gaps 2;
QY 137	gacagtatactatgcattagggcgcaaacaccacctccgtctctgatgggtgagatgg 196
DB 101	GGAAGGACTTTGCCATAGGTGGCTGAGCGCCACCATCTCTCTTACTTGCCCAAGGGCG 160
QY 197	gaaaaaaaagatagttccttcatttgcgtataagaagacgacgacgacgccattgg 256
DB 161	TAAAAAGATAG-----TCCTCCCATTAGCTAGAGAGCAAAACCCC-AGAAAGCCTATTGG 213
QY 257	ttagtgccgcggcgccgttgtcttcgcaagccgctagaggctaccggcgagggg 316
DB 214	CTGCGCGCTCGCGGGGCTTGGTCCGCTTTGAAGCGGGCTCGCGCTCGAGAGAGGGT 273
QY 317	cggcgccgagctgcgcgttcggtttaccagagacacgtgcgcgaagcgg 376
DB 274	GGCGGGAGGCTAGCTTGTCTGTGGTTGTCGAGGACACGTGTGACGTCCCGGAAGCGG 333
QY 377	ccgggggaagctgctccgcgcgcgtgcgcgaggaagaccccgggtccgctctgctc 436
DB 334	CGAGGGGAAACTGCTCCGCGCGCGCGGGAGGAGAACCCCGCCGGTCCITTAGGGTCC 393
QY 437	gggtccgctgggccaatggagtcattctctgagc 470
DB 394	GGGCCCCGGCGGCCCATGGATTCAATGCTTGACC 427
RESULT	7
LOCUS	BE908458 848 bp mRNA linear EST 20-OCT-2000

BC011767
VERSION
KEYWORDS
SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (30-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cdNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center (NISC),
Gaithersburg, Maryland:
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karling,E., Legaspi,R.,
Lim,W., Maduro,Q.L., Mastello,C., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Snyder,B., Stantirpop,S., Thomas,P.J.,
Tiongson,E.E., Touchman,J.W., Tsurgueon,C., Vogt,J.L., Walker,M.A.,
Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 27 Row: F Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein
This clone has the following problem: incomplete processing.

FEATURES
source
Location/Qualifiers
1..1602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3609644"
/tissue_type="Uterus, endometrium adenocarcinoma"
/clone_lib="NIH_MGC_44"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

BASE COUNT 361 a 452 c 491 g 288 t
ORIGIN

Query Match	33.9%	Score 159.8;	DB 11;	Length 1602;
Best Local Similarity	71.6%;	Pred. No. 2.9e-25;		
Matches 239;	Conservative	0;	Mismatches 88;	Indels 7;
Gaps				

```

QY 137 ggcaagtacgtataggcgccaaagccaccatcgctctctgattgggtgagatgg 196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 GGAAAGGACATTGGCATAGTCGTGTAGGCCACCATCTGCTCTTACTTGCCCAAGGCGC 140
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 197 gaaaaaaaagaatgctcctctatggcttaaaagcacgcgcgagcaaccatgg 256
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 TAAAAAGATAG-----TCCTCCCATTAGCTAGAGAGCAAACCCC-AGAAAGCCTATTGG 193
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 257 ttngtcgcccgcggttggttcggtttcgcaagcgcgctagagctaccggcgagggg 316
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 CTGCGCGCTCGCGGGCCTTGGTCCGCTTTTGAAGCGGGCTTCGCGCTCGAGAGGAGGCGC 253
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 317 cgggccggagctgcgcttgcgttggttaccagagacagctgcgtagtcccgaagcgg 376
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 GGGCGGAGGAGCTAGCTGTGTGCTGGTTCGTTCGGAGGACACGTGTGTCATCCCGAAGCGG 313
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

QY	257	ttngtgcgccgagggccttggtgctggttcgaagccgcgtagaggctaccgcggcgagggg	316
Db	153	CTGCGCCGTCCGGGGGCTTTGGTCCGCTTTGAAGCGGGGTGCAGTCGAGAGAGGAGGC	212
QY	317	cgggcggagaactgcgcgttccgttggtttaccagagacagctgcgcagtcccgaagacgg	376
Db	213	GGCGGGGAGGCTAGCTGTTGTCTGGTGTTGCTCGGAGGACACTGTGCAGTCCCGGAAGCGG	272
QY	377	ccggggaaagtgcctgcgcgcgcgtccgagagaaagccgcgcgcgtctgcctctgctct	436
Db	273	CGAGGGAAACTGCTCCGCGCGCGCGCGGAGGAGAACGCCCGCTCTTTAGGTCC	332
QY	437	gggtccggtcgggccaatggagtcgatctctgagc	470
Db	333	GGGCGCGCGGGCATGGATTCAATGCCTGAGC	366

RESULT 12

BI766238
LOCUS

DEFINITION BI766238 Homo sapiens cdna clone IMAGE:520525 5', mRNA sequence.

ACCESSION BI766238

VERSION BI766238.1 GI:15757816

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS 1 (bases 1 to 706)

TITLE NIH-MGC http://mgc.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-f@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1507 row: 1 column: 22
High quality sequence stop: 706.

FEATURES
source Location/Qualifiers
1..706
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:520525"
 /clone_lib="NIH_MGC_122"
 /lab_host="DH10B"
 /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."

BASE COUNT 130 a 203 c 233 g 139 t 1 others

ORIGIN

Query Match 33.6%; Score 158.2; DB 10; Length 706;

Best Local Similarity 71.3%; Pred. No. 6.4e-25;

Matches 238; Conservative 0; Mismatches 89; Indels 7; Gaps 2;

QY 137 ggcaagtcatcgcctagggcgcaaagccaccatccgctctctgattgggtgagatcg 196

Db 76 GGAAGAAGCACTTGGCATAGTGCCTGAGGCCACCCTGCTCTCTACTGCCAAGGCGC 135

```

QY 197 gaaaaaaagatagttctctcattggtataaagcagacccgagcgaacccattgg 256
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 136 TAAAAAGATAG-----TCTTCCCATAGCTAGAGAGAAAC-CCAGAAAGCCTATTGG 188
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 257 ttngtgcgcgcgcgcgccttgctggtttgcgaagcctagagctaccgagcgagggg 316
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 189 CTGCGCGCTCCGCGGCGCTTGTGTCCTTTGAAGCGGCGTGCCTGCGAGAGAGGGC 248
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 317 cgggcgcgcgcgcgccttgctggtttaccagagacacgtgcgcagtcgcgcgaagcgg 376
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 249 GGGCGGAGGCTAGCTGTGCTGCTGCTGCGAGCAGCTGTGCAGTCCCGGAAGCGG 308
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 377 cggggggaagctgctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 436
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 309 CGAGGGGAAACTGCTCCGCGCGCGCGCGGAGGAGGAACCGCGCGTCTCTTTAGGTC 368
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 437 ggtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 470
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 369 GGGCGCGCGCGCGCGCGCTGATTCATGCTGAGC 402
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
BG827636 602748652F1 NIH_MGC_17 827 bp mRNA linear EST 22-MAY-2001
LOCUS 602748652F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4901437 5',
DEFINITION mRNA sequence.
ACCESSION BG827636
VERSION BG827636.1 GI:14175223
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 827)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1796 row: h column: 14
High quality sequence stop: 535.
FEATURES
Location/Qualifiers
1..827
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4901437"
/tissue_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT
ORIGIN

Query Match 33.6%; Score 158.2; DB 10; Length 827;
Best Local Similarity 71.3%; Pred. No. 6.4e-25;
Matches 238; Conservative 0; Mismatches 89; Indels 7; Gaps 2;

```

QY 137 ggcagtgcctataggtgagcgaagccaccaccatccgctctctctgattgggtgagatgg 196
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 GGAAGAGACTTTGCCATAGTCTGCTGAGGCCACCATCTGCTCTTACTGGCCAAGGCG 126
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 197 gaaaaaaagatagttctctcattggtataaagcagacccgagcgaacccattgg 256
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 TAAAAAGATAG-----TCTTCCCATAGCTAGAGAGCAACCCC-AGAAAGCCTATTGG 179
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 257 ttngtgcgcgcgcgcgccttgctggtttgcgaagcctagagctaccgagcgagggg 316
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 180 CTGCGCGCTCCGCGGCGCTTGTGTCCTTTGAAGCGGCGCTGCGGCTGCGAGGAGGGC 239
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 317 cgggcgcgcgcgcgccttgctggtttaccagagacacgtgcgcagtcgcgcgcgcgcgcgc 376
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 240 GGGCGGAGGCTAGCTGCTCCGAGCGCGCGCGGAGGAGGAACCGCGCGTCTCTTTAGGTC 299
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 377 cggggggaagctgctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 436
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 300 CGAGGGGAAACTGCTCCGAGCGCGCGCGGAGGAGGAACCGCGCGTCTCTTTAGGTC 359
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 437 ggtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 470
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 360 GGGCGCGCGCGCGCGCATGATTCATGCTGAGC 393
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
BG755172 602711494F1 NIH_MGC_48 951 bp mRNA linear EST 15-MAY-2001
LOCUS 602711494F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851982 5',
DEFINITION mRNA sequence.
ACCESSION BG755172
VERSION BG755172.1 GI:14065825
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 951)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1695 row: k column: 23
High quality sequence stop: 831.
FEATURES
Location/Qualifiers
1..951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4851982"
/tissue_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT
ORIGIN

Query Match 33.6%; Score 158.2; DB 10; Length 951;
Best Local Similarity 71.3%; Pred. No. 6.4e-25;
Matches 238; Conservative 0; Mismatches 89; Indels 7; Gaps 2;

```
Oy 137 ggcagtgatatacgctatagggcgcaaaagccaccatccgctctctctgattgggtgagatgg 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 82 GGAAGAGACTTTGCCATAGGTCGGTAGGCCACCACCATCTGCTCTTACTGGCCAAAGGCG 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 197 gaaaaaaagatagttccctctctattgctataaaagcagacgcgcgagcgaacccattgg 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 TAAAAAGATAG-----TCTTCCCATTAGCTAGAGAGCAAAAC-CCAGAAAAGCCCTATTGG 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 257 ttgntgcgcgcgcgcgcgttgcgttttcgcaagccctagaggtaccgcgggcgagggg 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 CTGCGCCGTCGCGGGCCCTTGGTCGGCTTTGAAGCGGGCTGCGGCTCGAGAGAGGCGC 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 317 cgggcggagctgcgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 GGGCGGAGGCTAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 377 cggggggaagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 CGAGGGGAAGCTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 437 ggggtccggtggcccatgagtcctgagc 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

BI523540 603175461F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5239791 5',
LOCUS mRNA sequence.

ACCESSION BI523540.1 GI:15348332

VERSION EST.

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 724)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1A11604 row: j column: 16

High quality sequence start: 3

High quality sequence stop: 724.

Location/Qualifiers

FEATURES

source

1..724

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5239791"

/clone_lib="NIH_MGC_121"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: EcoRV (destroyed); RNA source anonymous pool of 3

fetal brains, female age 20 weeks, female age 24 weeks,

and male age 26 weeks. Library is oligo-dT primed and

directionally cloned (EcoRV site is destroyed upon

cloning). Average insert size 1.7 kb, insert size range

0.7-3.5 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 017. Note:

BASE COUNT 134 a 203 c 241 g 146 t
ORIGIN

Query Match 33.2%; Score 156.2; DB 10; Length 724;

Best Local Similarity 71.9%; Pred. No. 1.8e-24;

Matches 233; Conservative 0; Mismatches 84; Indels 7; Gaps 2;

```
Oy 147 tcgctatagggcgcaaaagccaccatccgctctctctgattgggtgagatgggaaaaaaa 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 TTGCCATAGGTCGGTAGGCCACCACCATCTGCTCTTACTGGCCAAAGGCGCTAAAAAGATA 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 207 agatagttccctctattgctataaaagcagacgcgcgagcgaacccattggttngtcgcc 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 -----GTCCTTCCCATTAGCTAGAGAGCAAAAC-CCAGAAAAGCCCTATTGGCTGCCCGCTC 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 267 cgcggccttgggtcggttttcgcaagccgctagaggctaccggtccgagggcggcgcggag 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 CGCGGGCCCTTGGTCGGCTTTGAAGCGGGCTGCGGCTCGAGAGAGGCGGCGCGGAGG 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 327 ctgcgcgttgcgttgggttaccacagagacacgtgcgcagtcgccggaagcgccgggggaag 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 CTAGCTGTGTCGTTGCTCGGAGGACGCTGTCAGTCCCGGAAGCGCGGCGGAGGAGAA 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 387 ctgctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 CTGCTCCCGCGCGCGCGCGCGGAGGAGAAACCGCGCGTCTTTAGGGTCCCGGGCCCGGCC 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 447 gggccatggagtcctatgctgagc 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GGGCCATGGATTCAATGCTGAGC 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 23, 2002, 23:06:04

Job time: 7520 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:20:16 ; Search time 121.39 Seconds
(without alignments)
953.072 Million cell updates/sec

Title: US-09-696-686-50

Perfect score: 471
Sequence: 1 acagaaacagaacaaaa.....atggagtccatgtctgagct 471

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/1/lna/5A-COMB.seq: *
2: /cgn2_6/ptodata/1/lna/5B-COMB.seq: *
3: /cgn2_6/ptodata/1/lna/6A-COMB.seq: *
4: /cgn2_6/ptodata/1/lna/6B-COMB.seq: *
5: /cgn2_6/ptodata/1/lna/PCTUS-COMB.seq: *
6: /cgn2_6/ptodata/1/lna/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.6	8.2	5539	4	US-08-628-829-3
2	38.4	8.2	5253	4	US-09-423-890-7
3	36.8	7.8	5011	1	US-08-141-893-1
4	36.8	7.8	5011	1	US-08-463-092B-1
5	36.8	7.8	5011	1	US-08-463-092B-3
6	36.8	7.8	5011	2	US-08-462-109A-1
7	36.8	7.8	5011	2	US-08-462-109A-3
8	36.8	7.8	5011	2	US-08-460-907B-1
9	36.8	7.8	5011	2	US-08-460-907B-3
10	36.8	7.8	5011	3	US-08-463-179A-1
11	36.8	7.8	5011	3	US-08-463-179A-3
12	36.8	7.8	5011	3	US-08-461-384B-1
13	36.8	7.8	5011	3	US-08-461-384B-3
14	36.8	7.8	5011	3	US-08-407-207A-1
15	36	7.6	1120	3	US-08-586-165-1
16	36	7.6	3937	3	US-08-586-165-8
17	35	7.4	1053	6	5352575-6
18	34.6	7.3	6453	1	US-08-306-691B-14
19	34.6	7.3	6453	1	US-09-209-668-10
20	34.6	7.3	6453	3	US-09-356-952-8
21	34	7.2	23673	4	US-09-773-816-1
22	34	7.2	71989	4	US-09-443-501A-2
23	33.8	7.2	43280	2	US-08-804-227C-1
24	33.6	7.1	15079	4	US-09-385-028-1
25	33.6	7.1	4403765	4	US-09-103-840A-2
26	33.4	7.1	888	4	US-08-765-907A-2
27	33.4	7.1	2888	4	US-08-765-907A-1

28	33.4	7.1	3107	2	US-08-813-940-3	Sequence 3, Appli
29	33.2	7.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
30	33	7.0	1224	1	US-08-924-847A-1	Sequence 1, Appli
31	33	7.0	1224	1	US-08-924-847A-3	Sequence 3, Appli
32	33	7.0	1224	3	US-09-120-052-1	Sequence 1, Appli
33	33	7.0	1224	3	US-09-120-052-3	Sequence 3, Appli
34	33	7.0	1428	4	US-09-118-442-5	Sequence 5, Appli
35	33	7.0	1428	4	US-09-677-064-5	Sequence 5, Appli
36	32.8	7.0	985	4	US-09-056-556-182	Sequence 18, App
37	32.8	7.0	4481	4	US-09-041-886-18	Sequence 18, App
38	32.8	7.0	5228	4	US-09-428-711A-15	Sequence 15, Appl
39	32.8	7.0	6244	1	US-08-078-726-15	Sequence 15, Appl
40	32.8	7.0	6244	1	US-08-260-452-8	Sequence 8, Appli
41	32.8	7.0	6244	2	US-08-481-970-8	Sequence 8, Appli
42	32.8	7.0	6244	2	US-08-897-719-8	Sequence 8, Appli
43	32.8	7.0	6244	4	US-09-163-269-8	Sequence 8, Appli
44	32.8	7.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
45	32.4	6.9	626	2	US-08-696-376-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-628-829-3
; Sequence 3, Application US/08628829A
; Patent No. 6333170

; GENERAL INFORMATION:
; APPLICANT: Johnson, Gary L.

; TITLE OF INVENTION: Method And Product For Regulating Cell Responsiveness To Enter

; FILE REFERENCE: CPI-004DVCP3

; CURRENT APPLICATION NUMBER: US/08/628,829A

; CURRENT FILING DATE: 1996-04-05

; EARLIER APPLICATION NUMBER: 08/440,421

; EARLIER FILING DATE: 1995-05-15

; EARLIER APPLICATION NUMBER: 08/323,460

; EARLIER FILING DATE: 1994-10-14

; EARLIER APPLICATION NUMBER: 08/049,254

; EARLIER FILING DATE: 1993-05-15

; EARLIER APPLICATION NUMBER: 08/410,602

; EARLIER FILING DATE: 1995-04-24

; EARLIER APPLICATION NUMBER: 08/472,934

; EARLIER FILING DATE: 1995-06-06

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 5539

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(4779)

US-08-628-829-3

Query Match 8.2% Score 38.6; DB 4; Length 5539;
Best Local Similarity 55.6% Pred. No. 0.067;

Matches 74; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy	287	gcaagcgcctagagctaccgagcggcgagggcgccgagctcgccgtggttac	346
Db	287	gtagcccgagagaaaaatggcgcgcgcgagctcgctcgctcgctcgcttc	346
Qy	347	ccagagacagctgcgcagctcccgagagcggcgccgagctcgctcgcttc	406
Db	347	cgggcgcgcgcgcgagctcccgagagcggcgccgagctcgctcgcttc	406
Qy	407	gaggaagcgcgcg 419	
Db	407	agggaaagcgcgcg 419	

RESULT 2

[illegible]

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/407,207
;; FILING DATE: 20-MAR-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Steeg, Carol Miernicki
;; REGISTRATION NUMBER: 39,539
;; REFERENCE/DOCKET NUMBER: Q1546
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (613) 545-2342
;; TELEFAX: (613) 545-6853
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5011 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 196..4788
US-08-463-092B-1

Query Match 7.8%; Score 36.8; DB 1; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 309 gcgagggcgggcgagctgcggtggttaccagagacagtcgagtgccc 368
DB 114 GCCCGGTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 173
QY 369 ggaagcgggggggaagctgctgcgaggaagcgagcgagcggtccgc 428
DB 174 CGTGC 233
QY 429 tctgctctggtccgg 444
DB 234 CCCGCTCTGGGACTGG 249

RESULT 5
US-08-463-092B-3
; Sequence 3, Application US/08463092B
; Patent No. 5766880
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,092B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,923
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/029,340

;; FILING DATE: 8-MAR-1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/141,893
;; FILING DATE: 26-OCT-1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/407,207
;; FILING DATE: 20-MAR-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Steeg, Carol Miernicki
;; REGISTRATION NUMBER: 39,539
;; REFERENCE/DOCKET NUMBER: Q1546
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (613) 545-2342
;; TELEFAX: (613) 545-6853
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5011 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 196..4788
US-08-463-092B-3
Query Match 7.8%; Score 36.8; DB 1; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 309 gcgagggcgggcgagctgcggtggttaccagagacagtcgagtgccc 368
DB 114 GCCCGGTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 173
QY 369 ggaagcgggggggaagctgctgcgaggaagcgagcgagcggtccgc 428
DB 174 CGTGC 233
QY 429 tctgctctggtccgg 444
DB 234 CCCGCTCTGGGACTGG 249

RESULT 6
US-08-462-109A-1
; Sequence 1, Application US/08462109A
; Patent No. 5882875
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: MULTIDRUG RESISTANT TUMOR CELLS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,109A
; FILING DATE:
; CLASSIFICATION: 435

RESULT 7
US-08-462-109A-3
Sequence 3, Application US/08462109A
Patent No. 5882875
GENERAL INFORMATION:
APPLICANT: Cole, Susan P.C.
APPLICANT: Deesley, Roger G.
TITLE OF INVENTION: METHODS FOR IDENTIFYING
TITLE OF INVENTION: MULTIDRUG RESISTANT TUMOR CELLS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,109A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

```

RESULT      8
US-08-460-907B-1
; Sequence 1, Application US/08460907B
; Patent No. 5891724
; GENERAL INFORMATION:
;   APPLICANT: Deeley, Roger G.
;   APPLICANT: Cole, Susan P.C.
;   TITLE OF INVENTION: METHODS FOR CONFERRING MULTIDRUG
;   TITLE OF INVENTION: RESISTANCE ON A CELL
;   NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
;   STREET: Queen's University at Kingston
;   CITY: Kingston
;   STATE: Ontario
;   COUNTRY: CANADA
;   ZIP: K7L 3N6
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: ASCII text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/460,907B
;   FILING DATE: 05-JUN-1995
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   PRIOR APPLICATION NUMBER: 07/966,923

```



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,179A
FILING DATE:
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A. Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: POI-002CP8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5011 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..4788
US-08-463-179A-1

Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 309 gcagagggcgccgagctccgttgcgttggtaccagagacacgtgcgagtcgcc 368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 GCCCGGTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173
QY 369 ggaagcgccgggggagctgcgcgtccgcgcgcgcgcgcgcgcgcgcgcgc 428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233
QY 429 tctgtctctgggtccgg 444
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 CCGCGCTCTGGGACTGG 249

RESULT 11
US-08-463-179A-3
; Sequence 3, Application US/08463179A
; Patent No. 6001563
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING CHEMOSENSITIZERS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM: ...

COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,179A
FILING DATE:
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A. Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: POI-002CP8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5011 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..4788
US-08-463-179A-1

Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 309 gcagagggcgccgagctccgttgcgttggtaccagagacacgtgcgagtcgcc 368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 GCCCGGTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173
QY 369 ggaagcgccgggggagctgcgcgtccgcgcgcgcgcgcgcgcgcgcgcgc 428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233
QY 429 tctgtctctgggtccgg 444
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 CCGCGCTCTGGGACTGG 249

RESULT 12
US-08-461-384B-1
; Sequence 1, Application US/08461384B
; Patent No. 6025473
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```
;; SOFTWARE: ASCII text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/461,384B
;; FILING DATE: 05-JUN-95
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/966,923
;; FILING DATE: 27-OCT-1992
;; APPLICATION NUMBER: 08/029,340
;; FILING DATE: 8-MAR-1993
;; APPLICATION NUMBER: 08/141,893
;; FILING DATE: 26-OCT-1993
;; APPLICATION NUMBER: 08/407,207
;; FILING DATE: 20-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Steeg, Carol Miernicki
;; REGISTRATION NUMBER: 39,539
;; REFERENCE/DOCKET NUMBER: Q1547
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (613) 545-2342
;; TELEFAX: (613) 545-6853
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5011 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 196..4788
US-08-461-384B-1
```

```
Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 309 gcaaggggcgagagagtcggttccggtggtaccagagacagtcgcagtgccc 368
Db 114 GCCCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173

Qy 369 gaaagcgccgggggaaagtcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
Db 174 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233

Qy 429 tctgctctgggtccgg 444
Db 234 CCCGCTCTGGGACTGG 249
```

```
RESULT 13
US-08-461-384B-3
; Sequence 3, Application US/08461384B
; Patent No. 6025473
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,384B
```

```
;; FILING DATE: 05-JUN-95
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/966,923
;; FILING DATE: 27-OCT-1992
;; APPLICATION NUMBER: 08/029,340
;; FILING DATE: 8-MAR-1993
;; APPLICATION NUMBER: 08/141,893
;; FILING DATE: 26-OCT-1993
;; APPLICATION NUMBER: 08/407,207
;; FILING DATE: 20-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Steeg, Carol Miernicki
;; REGISTRATION NUMBER: 39,539
;; REFERENCE/DOCKET NUMBER: Q1547
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (613) 545-2342
;; TELEFAX: (613) 545-6853
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5011 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 196..4788
US-08-461-384B-3
```

```
Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 309 gcaaggggcgagagagtcggttccggtggtaccagagacagtcgcagtgccc 368
Db 114 GCCCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173

Qy 369 gaaagcgccgggggaaagtcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
Db 174 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233

Qy 429 tctgctctgggtccgg 444
Db 234 CCCGCTCTGGGACTGG 249
```

```
RESULT 14
US-08-407-207A-1
; Sequence 1, Application US/08407207A
; Patent No. 6063621
; GENERAL INFORMATION:
; APPLICANT: Deeley, Roger G.
; APPLICANT: Cole, Susan P.C.
; TITLE OF INVENTION: ANTIBODIES TO A MULTIDRUG RESISTANCE PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/407,207A
; FILING DATE: 20-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Steeg, Carol Miernicki
REGISTRATION NUMBER: 39,539
REFERENCE/DOCKET NUMBER: Q1512
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 545-2342
TELEFAX: (613) 545-6853
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..4788
US-08-407-207A-1

Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 309 gcaggggcccgcgagctccgttgcgttaccagagacagtgccgagtcgcc 368
Db 114 GCCCGGTGCCCCGCGCCCGCCGCGCAGCAACCGGCGCGATCACCCGCGCGCGTGCC 173
QY 369 ggaagcggcgggggaagtgtctccgcgcgtccgagggaagcgcgcgggtccgc 428
Db 174 CGCCGCGCGCGCGCCACCGGATGCGGTCCGGGCTTCGAGCGCGGATGGCTCCGA 233
QY 429 tctgtctgggtccgg 444
Db 234 CCCGCTCTGGGACTGG 249

RESULT 15

US-08-586-165-1
Sequence 1, Application US/08586165
Patent No. 6054298

GENERAL INFORMATION:

APPLICANT: Laufer, Edward M.
APPLICANT: Orozco, Olivia E.
APPLICANT: Tabin, Clifford J.
TITLE OF INVENTION: Fringe Proteins and Pattern Formation
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/586,165
FILING DATE: 16-JAN-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: HU95-05

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1120 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..240, 244..474, 478..531, 538..579, 583
LOCATION: ..678, 682..687, 691..807, 811..843, 850..870,
LOCATION: 874..990, 994..1056, 1060..1083, 1087..1104, 1108
LOCATION: ..1119)
US-08-586-165-1

Query Match 7.6%; Score 36; DB 3; Length 1120;
Best Local Similarity 50.6%; Pred. No. 0.2;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 284 ttcgcaagccgctagaggctaccggcgaggcgggcgagcgtccggttgcctggt 343
Db 65 TTCTCTGCTGCTGCTGCGCCCGGGGACAGCCCCCGCGCGCGCGCGCGCGCG 124
QY 344 taccagagacacgtgcagtcctccggaagcgcgcggggaagctcctccgcgcgtg 403
Db 125 CCCGCGGCGCCACGAGGCCCTCCCCGAAAGCGGAGCGCGCGCGCGCGCGCG 184
QY 404 ccggagggaagcgcgcgggtccgtctctctgtggtccggtgggccaatgg 455
Db 185 CCCGGGACCCGCGCGCGGCTCGGAGCGCGGGGGCGCGCGCGCGCGCGCG 236

Search completed: May 24, 2002, 00:20:45
Job time: 11711 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 24, 2002, 00:23:41 ; Search time 525.6 Seconds
(without alignments)
1538.559 Million cell updates/sec

Title: US-09-696-686-50

Perfect score: 471

Sequence: 1 acagaaaaacaagaacacaaaa.....atggagtcacatgtctgagct 471

Scoring table: IDENTITY_NUC

Gap 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	470	99.8	471	22	AAS05284
2	158.2	33.6	1675	22	T243 gene homologous sequence
3	148	31.4	1657	20	Human CDNA sequenc
4	66	14.0	1839	22	Human secreted pro
5	66	14.0	1848	22	Murine trinucleoti
6	46.8	9.9	2829	22	DNA sequence of an
7	39.6	8.4	48037	21	Human OREF ORF790
8	39.6	8.4	48037	22	Human immune/haema
9	39.6	8.4	48045	22	Human immune/haema

C 10	39.6	8.4	48045	22	AAK85984	Human immune/haema
C 11	39.2	8.3	527	21	AAZ93044	Human pancreatic c
C 12	39	8.3	1845	21	AAZ98400	Canine beta1-adren
C 13	38.6	8.2	3559	22	AAI60356	Human polynucleoti
C 14	38.6	8.2	3577	22	AAI58570	Human polynucleoti
C 15	38.6	8.2	3623	23	AAAS4706	DNA encoding novel
C 16	38.4	8.2	1614	23	AAAH26499	Human low density
C 17	38.4	8.2	5253	20	AAZ25072	Murine MEKK1 nucle
C 18	38.4	8.2	5253	20	AAAX0911	Murine MEKK1 CDNA
C 19	38.4	8.2	12425	22	AAAH26495	Human low density
C 20	38.2	8.1	293	20	AAZ32038	Human METH2 relate
C 21	38.2	8.1	293	22	AAC90095	HTFCC34RA CDNA clo
C 22	38.2	8.1	18660	21	AAAS8472	Nucleotide sequenc
C 23	38	8.1	3544	23	AAAS81565	DNA encoding novel
C 24	37.8	8.0	2346	22	AAK52993	Human polynucleoti
C 25	37.8	8.0	2710	22	AAK52009	Human polynucleoti
C 26	37.6	8.0	5131	22	AAAL37435	Human musculoskele
C 27	37.6	8.0	5131	22	AAAL37435	Human immune/haema
C 28	37.6	8.0	5139	22	AAAL37437	Human immune/haema
C 29	37.6	8.0	5139	22	AAAL37437	Human musculoskele
C 30	37	7.9	1457	22	AAI58521	Human immune/haema
C 31	36.8	7.8	4423	17	AAAI4914	Human polynucleoti
C 32	36.8	7.8	4669	17	AAAI4913	CDNA encoding mult
C 33	36.8	7.8	4834	17	AAAI4912	CDNA encoding mult
C 34	36.8	7.8	4864	17	AAAI4911	CDNA encoding mult
C 35	36.8	7.8	4885	17	AAAI4910	CDNA encoding mult
C 36	36.8	7.8	5011	15	AAO65377	CDNA encoding mult
C 37	36.8	7.8	5011	17	AAAI17173	Multidrug resistanc
C 38	36.8	7.8	5011	19	AAV311497	CDNA encoding mult
C 39	36.8	7.8	5011	19	AAV311498	Human multidrug re
C 40	36.8	7.8	5011	20	AAAX19817	Human MRP variant
C 41	36.8	7.8	5011	20	AAAX19818	Human multidrug re
C 42	36.8	7.8	5011	20	AAAX19817	Human MDR variant
C 43	36.8	7.8	5011	21	AAAX21976	Human multidrug re
C 44	36.8	7.8	5011	21	AAAX21976	Multidrug resistanc
C 45	36.8	7.8	5011	21	AAAX21976	Human ATP binding

ALIGNMENTS

RESULT 1
AAS05284
ID AAS05284 standard; DNA; 471 BP.
XX
AC AAS05284;
XX
DT 07-SEP-2001 (first entry)
XX
DE T243 gene homologous sequence #1 generated by PCR.
XX
KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; ds.
XX
OS Mus sp.
XX
PN WO200130798-A1.
XX
PD 03-MAY-2001.
XX
PF 26-OCT-2000; ~2000WO-US29382.
XX
PR 26-OCT-1999; 99US-0161488.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Klein R, Matthews W, Moore M, Allen KD;
XX
DR WPI; 2001-300473/31.
XX
PT Novel transgenic animals useful as animal model for characterization of
function of a gene encoding trinucleotide repeat proteins (TRPs),

Fri May 24 10:49:47 2002

us-09-696-686-50.rng

PT contains heterozygous disruption in a gene encoding TRP -

XX Claim 11; Fig 14; 106pp; English.

XX The present sequence for T243 homologous sequence #1 is generated by PCR
CC and is homologous to the T243 gene which encodes for a trinucleotide
CC repeat protein (TRP). The invention describes methods of producing
CC embryonic stem (ES) cells comprising a heterozygous disruption in a
CC target DNA sequence (preferably gene T243) encoding a TRP and of
CC producing a knockout mouse comprising a homozygous disruption in a gene
CC encoding TRP, where the disruption inhibits the production of the wild
CC type TRP. The invention also relates to identifying agents capable of
CC affecting a phenotype of a knockout mouse. Also described are methods of
CC determining whether expansion of the trinucleotide repeat in a gene
CC encoding TRP produces a phenotypic change. The transgenic animals and the
CC cells are useful for identifying compounds capable of ameliorating
CC disease symptoms, and as test substrates for the identification of drugs,
CC pharmaceuticals, therapies and interventions which may be effective in
CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
CC Huntington's disease. The animal models for trinucleotide repeat
CC disorders are ideal model systems to study the progression of disease in
CC vivo, the molecular basis of these diseases and show the features
CC observed in human disease. Using the mice, it is possible to model both
CC the pathogenic mechanism and the trinucleotide repeat instability in the
CC mouse.

XX Sequence 471 BP; 105 A; 129 C; 148 G; 88 T; 1 other;

Query Match 99.8%; Score 470; DB 22; Length 471;
Best Local Similarity 100.0%; Pred. No. 7e-125;
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acagaaacacgaagaaacacacatgaagatagctgttattccacgggctagatgaccca 60
Db 1 acagaaacacgaagaaacacacatgaagatagctgttattccacgggctagatgaccca 60
Qy 61 aggtggttccatcaagtgatgaaggttcacccgctagggaactgactccagctac 120
Db 61 aggtggttccatcaagtgatgaaggttcacccgctagggaactgactccagctac 120
Qy 121 tgagctctcttagctgagctgatatcgctatagggcgccaaagccaccatcgctctc 180
Db 121 tgagctctcttagctgagctgatatcgctatagggcgccaaagccaccatcgctctc 180
Qy 181 tgattgggtgagtgaggaagaaagaaagatagttctctcattggtctataaagcagacgc 240
Db 181 tgattgggtgagtgaggaagaaagaaagatagttctctcattggtctataaagcagacgc 240
Qy 241 cgagcgaaacccattggttgnctgcgcgcgggcttgggtggttcgcaagcgcgtagag 300
Db 241 cgagcgaaacccattggttgnctgcgcgcgggcttgggtggttcgcaagcgcgtagag 300
Qy 301 gctaccggcgagggcgggcgagctgcgcgttgcgtgttaccagagacacgtgc 360
Db 301 gctaccggcgagggcgggcgagctgcgcgttgcgtgttaccagagacacacgtgc 360
Qy 361 cgagctccggaagcgcgcgcggggaagctgctccgcgcgcgtgcgaggaagcgcgcgc 420
Db 361 cgagctccggaagcgcgcgcggggaagctgctccgcgcgcgtgcgaggaagcgcgcgc 420
Qy 421 gggctccgctctgctggttccgctggtggccatggagtcacatgctgagct 471
Db 421 gggctccgctctgctggttccgctggtggccatggagtcacatgctgagct 471

RESULT 2
ID AAS21311
XX AAS21311 standard; cDNA; 1675 BP.
AC AAS21311;
XX 24-OCT-2001 (first entry)
DT

XX DE
XX KW
KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO20010466-A2.
XX
PD 07-JUN-2001.
XX
PF 01-DEC-2000; 2000WO-US32678.
XX
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 09-DEC-1999; 99US-0170262.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 30-DEC-1999; 99WO-US31243.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 10-NOV-2000; 2000WO-US30873.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WL, Zhang Z;
XX
XX WPI: 2001-408281/43.
DR P-PSDB; AAU12239.
XX
XX Isolated, secretory and transmembrane PRO polypeptide used to detect
PT other PRO polypeptides, link bioactive molecules to cells expressing
PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT lung, breast, prostate, cervical.
XX
XX Claim 3; Fig 135; 813pp; English.
XX
XX AAS21244-AAS21518 encode for novel human secretory and transmembrane
CC PRO polypeptides. The PRO polypeptides are useful to detect other
CC PRO polypeptides, to link bioactive molecules to cells expressing
CC PRO polypeptides, to modulate biological activities of cells expressing
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample.
CC Some of the 275 sequences are also useful to stimulate the release of
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC proliferation or differentiation of chondrocytes, the proliferation or
CC gene expression in pericyte cells, the release of proteoglycans from
CC cartilage, the proliferation of inner ear utricular supporting cells or
CC of T-lymphocytes, the release of a cytokine from peripheral blood

Human cDNA sequence encoding for PRO4409 polypeptide.

Human secretory and transmembrane; PRO; mammalian; cancer; lung;
breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
adipocyte; A-peptide; factor VIIA; gene therapy; ss.

Homo sapiens.

WO20010466-A2.

07-JUN-2001.

01-DEC-2000; 2000WO-US32678.

01-DEC-1999; 99WO-US28301.

01-DEC-1999; 99WO-US28634.

02-DEC-1999; 99WO-US28551.

02-DEC-1999; 99WO-US28564.

02-DEC-1999; 99WO-US28565.

09-DEC-1999; 99US-0170262.

16-DEC-1999; 99WO-US30095.

20-DEC-1999; 99WO-US30911.

20-DEC-1999; 99WO-US30999.

30-DEC-1999; 99WO-US31243.

06-JAN-2000; 2000WO-US00277.

06-JAN-2000; 2000WO-US00376.

11-FEB-2000; 2000WO-US03565.

18-FEB-2000; 2000WO-US04341.

18-FEB-2000; 2000WO-US04342.

22-FEB-2000; 2000WO-US04414.

24-FEB-2000; 2000WO-US04914.

24-FEB-2000; 2000WO-US05004.

01-MAR-2000; 2000WO-US05601.

20-MAR-2000; 2000WO-US07377.

21-MAR-2000; 2000WO-US07532.

30-MAR-2000; 2000WO-US08439.

17-MAY-2000; 2000WO-US13705.

22-MAY-2000; 2000WO-US14042.

30-MAY-2000; 2000WO-US14941.

02-JUN-2000; 2000WO-US15264.

10-NOV-2000; 2000WO-US30873.

(GETH) GENENTECH INC.

Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WL, Zhang Z;

WPI: 2001-408281/43.

P-PSDB; AAU12239.

Isolated, secretory and transmembrane PRO polypeptide used to detect
other PRO polypeptides, link bioactive molecules to cells expressing
PRO polypeptides, and detect the presence of mammalian tumours e.g.
lung, breast, prostate, cervical.

Claim 3; Fig 135; 813pp; English.

AAS21244-AAS21518 encode for novel human secretory and transmembrane
PRO polypeptides. The PRO polypeptides are useful to detect other
PRO polypeptides, to link bioactive molecules to cells expressing
PRO polypeptides, to modulate biological activities of cells expressing
PRO polypeptides, and to detect the presence of mammalian lung, colon,
PRO polypeptides, and to detect the presence of mammalian lung, colon,
breast, prostate, rectal, cervical or liver tumours by comparing PRO
polypeptide expression in a cell sample to that in a control sample.
Some of the 275 sequences are also useful to stimulate the release of
tumour necrosis factor-alpha (TNF-alpha) from human blood, the
proliferation or differentiation of chondrocytes, the proliferation or
gene expression in pericyte cells, the release of proteoglycans from
cartilage, the proliferation of inner ear utricular supporting cells or
of T-lymphocytes, the release of a cytokine from peripheral blood

CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VI/A. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 other;

Query Match 33.6%; Score 158.2; DB 22; Length 1675;
 Best Local Similarity 71.3%; Pred. No. 2.1e-35;
 Matches 238; Conservative 0; Mismatches 89; Indels 7; Gaps 2;
 QY 137 ggcagtgtatcgctataggcgcccaaaagccaccatccgctctctctgattgggtgagatgg 196
 DB 29 ggaagagacttgccataggctgcgtgagggccaccatctgctcttactgcccaggcg 88
 QY 137 gaaaaaaaagatagttctctctcatgtgctataaagcagacgcccagcgaaaccattgg 256
 DB 89 taataagatag-----cttcccatagctagagagcaaac-cccagaaagcctattgg 141
 QY 257 ttngtgcggcgccggttggtgcttgcgaagcgcctagaggtaccggtgagggg 316
 DB 142 ctgcgcctccgcggcgttgctgccttgaaggcggtgcggtcgagaggggc 201
 QY 317 cggccgagagctgcctgctgcgtgttaccagagacacgtgcgagctcccgaaagcg 376
 DB 202 gggcgggaggtagctgtgcgtggtgtgcgagcagcgtgtgcagtcgcggaagcg 261
 QY 377 ccgggggaagctgtccgcgcgcgtgcgaggaagacgcccgcgggtccgctctctct 436
 DB 262 cgaggggaactgtccgcgcgcgcgcggcgagggagggaacccgcggctcttggggccc 321
 QY 437 ggggtccgggtggccatggagtcctgtgagc 470
 DB 322 gggccggcgccggccatggattcaatgctgagc 355

RESULT 3

AAX04377

ID AAX04377 standard; DNA; 1657 BP.

AC AAX04377;

DT 13-APR-1999 (first entry)

XX Human secreted protein gene 67 clone HTOFC34.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX WO9856804-A1.

XX 17-DEC-1998.

XX 11-JUN-1998; 98WO-US12125.

XX 02-OCT-1997; 97US-0051060.

XX 13-JUN-1997; 97US-0049547.

XX 13-JUN-1997; 97US-0049548.

XX 13-JUN-1997; 97US-0049549.

XX 13-JUN-1997; 97US-0049550.

XX 13-JUN-1997; 97US-0049606.

PR 13-JUN-1997; 97US-0049607.
 PR 13-JUN-1997; 97US-0049608.
 PR 13-JUN-1997; 97US-0049609.
 PR 13-JUN-1997; 97US-0049610.
 PR 13-JUN-1997; 97US-0049611.
 PR 13-JUN-1997; 97US-0050566.
 PR 13-JUN-1997; 97US-0050901.
 PR 13-JUN-1997; 97US-0052989.
 PR 08-JUL-1997; 97US-0051919.
 PR 12-AUG-1997; 97US-0055984.
 PR 12-SEP-1997; 97US-0058665.
 PR 12-SEP-1997; 97US-0058668.
 PR 12-SEP-1997; 97US-0058669.
 PR 12-SEP-1997; 97US-0058750.
 PR 12-SEP-1997; 97US-0058971.
 PR 12-SEP-1997; 97US-0058972.
 PR 12-SEP-1997; 97US-0058975.
 PR 02-OCT-1997; 97US-0060834.
 PR 02-OCT-1997; 97US-0060841.
 PR 02-OCT-1997; 97US-0060844.
 PR 02-OCT-1997; 97US-0060865.
 PR 02-OCT-1997; 97US-0061059.
 XX

(HUMA-) HUMAN GENOME SCI INC.

Brewer LA, Ebner R, Ferrie AM, Peng P, Greene JM, Lafleur DW;

Moore PA, NI J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;

PI Yu GL;

XX WPI; 1999-080881/07.

DR P-PSDB; AAW78192.

XX New isolated human genes and the secreted polypeptides they encode -
 useful for diagnosis and treatment of e.g. cancers, neurological
 disorders, immune diseases, inflammation or blood disorders

Claim 1; Page 233; 380pp; English.

CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number, and the clone it is derived
 CC from, are detailed in the descriptor line. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. AAX04302) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 86 novel genes and their fragments (nucleic
 CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 86
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX04311 for described uses).

XX Sequence 1657 BP; 361 A; 490 C; 503 G; 297 T; 6 other;

Query Match 31.4%; Score 148; DB 20; Length 1657;

Best Local Similarity 71.0%; Pred. No. 1.8e-32;

Matches 237; Conservative 1; Mismatches 88; Indels 8; Gaps 3;

QY 137 ggcagtgtatcgctataggcgcccaaaagccaccatccgctctctctgattgggtgagatgg 196

DB 127 ggaagagacttgccataggctgcgtgagggccaccatctgctcttactgcccaggcg 186

QY 197 gaaaaaaaagatagttctctctcatgtgctataaagcagacgcccagcgaacccattgg 256

DB 187 taataagatag-----tcytcccatagctagagagcaaacccc-agaagcctattgg 239

QY 257 ttngtgcggcggtggtgcttgcgaagccgctagaggtaccggtcgaggggg 316

DB 240 ctgcgcgtctccggtggtgcttgcgntttgaagcggtgctgcgtgcgagaggggc 299

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226581.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234977.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Disclosure; SEQ ID NO 39541; 3071pp + Sequence Listing; English.
PS
XX

that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention.

Sequence 48037 BP; 12892 A; 9814 C; 10466 G; 14865 T; 0 other;

Query Match	8.4%	Score 39.6;	DB 22;	Length 48037;
Best Local Similarity	50.5%;	Pred. No. 0.61;		
Matches	96;	Conservative	0;	Mismatches 94;
				Indels 0;
				Gaps 0;
Qy	265	ccccgggcttgctgggttcgaagccgctagagctaccggcgagggcgccggcg 324		
Db	420	CCGCTGGGCTCGCGCGCCGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 361		
Qy	325	agctgcgcgttgccgtggttaccagagacacgtgcagtcgccggaagcgccgggga 384		
Db	360	GGCTTCATGCTCCGCTGCTCCACGACTCAGGCGCGCGAGTCGCTGCTGCTGCCGCT 301		
Qy	385	agctgcgcgcgtgcgctgcggaggaagcgccggcggttcgctctgctggttcgg 444		
Db	300	GCTGCTCCG 241		
Qy	445	ctgggccaatg 454		
Db	240	GGGGGCGCGTG 231		

RESULT	9
AAK84730/C	
ID	AAK84730 standard; DNA; 48045 BP.
XX	
AC	AAK84730;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39542.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.

20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
17-NOV-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
01-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.
Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -
Disclosure: SEQ ID NO 40795; 3071pp + Sequence Listing: English.
AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome

PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225113.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 39542; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/hematopoietic-related diseases, especially cancers and cancer metastases of hematopoietic-derived cells. AAK64703

DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #20510.
DE	
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
XX	

XX WO200175067-A2.
PN

PD 11-OCT-2001.

XX
XX

PR 23-AUG-2000; 200003 0045107.
XX

XX
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631

DR WPI; 2001-639362/73.

XX
PT
New isolated polynucleotide and encoded polypeptides, useful in

PT responsible for genetic disorders of other traits and to assess biodiversity - biodiversity -

PS Claim 1; SEQ ID No 20510; 103pp; English.

the invention relates to a method, (I) is useful as hybridisation probes, polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC and gene mapping, and in recombinant production of (11). The polynucleotides are also used in diagnostics as expressed sequence

to restore normal activity of (II) or to treat disease states involving (II) for generating antibodies against it, detecting

quantities of a food supplement. (II) and its binding partners are useful in mediating a food supplement. (II) and its binding partners are useful for treating

The polypeptide and polynucleotide sequences have applications in

responsible for genetic disorders or other traits to assess products and to produce other types of data and products dependent on DNA analysis.

diagnostic coding sequences of the invention.

specification, but was obtained in accordance with the
at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3623 BP; 909 A; 865 C; 851 G; 993 T; 3 Other;

Query Match	8.28; Score 38.6; DB 23; Length 3623;

Matches 92; conservative 0, mismatches 50; mismatches

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

QY 311 gagggcgggcgaggctcgccgttgccgtggttaccagagacacgtgcgagtcgccg

DB GTTGGGCTCCGGGTTCCTCGCAGCCCAACCCCCCAATTTT

24 326 CAGTACGACCGCTCCAGGAGGGCAGGTGGCGGGGGCTGCCCGCGGAGCCGACC

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:14:24 ; Search time 3953.4 Seconds
(without alignments)
2493.145 Million cell updates/sec

Title: US-09-696-686-50

Perfect score: 471

Sequence: 1 acagaaacacgaagaaacaaa.....atggagtcctgctgagct 471

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description

C	1	451.6	95.9	89855	2	AC074220	AC074220 Mus muscu
C	2	159.8	33.9	1720	9	BC004423	BC004423 Homo sapi
C	3	158.2	33.6	113109	9	HS075N16	AL035587 Human DNA
C	4	148.8	31.6	962	9	HS080744	UB0744 Homo sapien
C	5	98	20.8	1882	10	BC013549	AC013549 Mus muscu
C	6	97	20.6	1909	10	AF361644	AF361644 Mus muscu
C	7	62.4	13.2	1422	9	BC008898	BC008898 Homo sapi
C	8	62.4	13.2	1422	9	BC008961	BC008961 Homo sapi
C	9	47.6	10.1	60424	2	AC090906	AC090906 Homo sapi
C	10	46.8	9.9	2828	9	AF130356	AF130356 Homo sapi
C	11	46.8	9.9	2941	9	AF316597	AF316597 Homo sapi
C	12	44.4	9.4	230254	2	AC074163	AC074163 Mus muscu
C	13	44	9.3	36224	1	SCD78	AL034355 Streptomy
C	14	44	9.3	205914	2	AC027682	AC027682 Homo sapi
C	15	43.2	9.2	201324	2	AC079637	AC079637 Mus muscu
C	16	42.8	9.1	104943	2	AC103066	AC103066 Rattus no
C	17	42.8	9.1	167501	2	AC097333	AC097333 Pan trogl
C	18	42.8	9.1	171431	2	AC106453	AC106453 Rattus no
C	19	42.6	9.0	90248	2	AC108366	AC108366 Pan trogl
C	20	42.4	9.0	125600	2	AC097133	AC097133 Rattus no
C	21	42	8.9	52884	2	AC068018	AC068018 Homo sapi
C	22	42	8.9	64114	2	AC104971	AC104971 Homo sapi
C	23	42	8.9	70119	2	AC022593	AC022593 Homo sapi
C	24	42	8.9	126529	2	AC105894	AC105894 Rattus no
C	25	41.6	8.8	90801	2	AC097703	AC097703 Rattus no
C	26	41.6	8.8	197025	2	AC105642	AC105642 Rattus no
C	27	41.4	8.8	126059	8	AP003197	AP003197 Oryza sat
C	28	41.2	8.7	92305	2	AC008516	AC008516 Homo sapi
C	29	41.2	8.7	151903	2	AC024349	AC024349 Homo sapi
C	30	41.2	8.7	178845	2	AC010242	AC010242 Homo sapi
C	31	41.2	8.7	180713	2	AC106818	AC106818 Homo sapi
C	32	41	8.7	5180	10	RU048596	U048596 Rattus norv
C	33	41	8.7	47387	2	AC098024	AC098024 Rattus no
C	34	41	8.7	91501	2	AC024344	AC024344 Homo sapi
C	35	41	8.7	112341	2	AC106237	AC106237 Rattus no
C	36	40.6	8.6	94296	2	AL160261	AL160261 Homo sapi
C	37	40.4	8.6	95536	2	AC096296	AC096296 Rattus no
C	38	40.4	8.6	121307	2	AC094145	AC094145 Rattus no
C	39	40.4	8.6	133641	2	AC106487	AC106487 Rattus no
C	40	40.4	8.6	182622	9	AC090567	AC090567 Homo sapi
C	41	40.2	8.5	92858	2	AC024959	AC024959 Homo sapi
C	42	40	8.5	64650	2	AC068864	AC068864 Homo sapi
C	43	40	8.5	125020	9	AF429315	AF429315 Homo sapi
C	44	39.8	8.5	3770	9	AF153418	AF153418 Homo sapi
C	45	39.8	8.5	5582	9	AF153416	AF153416 Homo sapi

ALIGNMENTS

RESULT 1
AC074220/c

LOCUS

DEFINITION

AC074220

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

JOURNAL

Submitted (20-JUL-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA
On Apr 28, 2001 this sequence version replaced gi:11276104.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site:
http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mouse.html
Contact: gnktm@capcod.bwh.harvard.edu

COMMENT

-----Summary Statistics
Center project name: ADY
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 79355 at least Q20
*Consensus quality: 75365 at least Q30
*Consensus quality: 69022 at least Q40
**Estimated insert size: agarose-PP - N/A
Quality coverage: agarose-PP - N/A
Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 7962: contig of 7962 bp in length
* 7963 7982: gap of unknown length
* 7983 13924: contig of 5942 bp in length
* 13925 13944: gap of unknown length
* 13945 18990: contig of 5046 bp in length
* 18991 19010: gap of unknown length
* 19011 22501: contig of 3491 bp in length
* 22502 22521: gap of unknown length
* 22522 27077: contig of 4556 bp in length
* 27078 27097: gap of unknown length
* 27098 28819: contig of 1722 bp in length
* 28820 28839: gap of unknown length
* 28840 31224: contig of 2385 bp in length
* 31225 31244: gap of unknown length
* 31245 35093: contig of 3849 bp in length
* 35094 35113: gap of unknown length
* 35114 38611: contig of 3498 bp in length
* 38612 38631: gap of unknown length
* 38632 42281: contig of 3630 bp in length
* 42282 42282: gap of unknown length
* 42283 44130: contig of 1849 bp in length
* 44131 44150: gap of unknown length
* 44151 47059: contig of 2909 bp in length
* 47060 47079: gap of unknown length
* 47080 49496: contig of 2417 bp in length
* 49497 49516: gap of unknown length
* 49517 52236: contig of 2720 bp in length
* 52237 52257: gap of unknown length
* 52258 54270: contig of 2014 bp in length
* 54271 54290: gap of unknown length
* 54291 56221: contig of 1981 bp in length
* 56222 56231: gap of unknown length
* 56232 58849: contig of 2558 bp in length
* 58850 58869: gap of unknown length
* 58870 60621: contig of 1752 bp in length
* 60622 60641: gap of unknown length
* 60642 62255: contig of 1614 bp in length
* 62256 62275: gap of unknown length
* 63085: contig of 810 bp in length
* 63086 63105: gap of unknown length
* 63106 65440: contig of 2335 bp in length
* 65441 65460: gap of unknown length

FEATURES
source

65461 67629: contig of 2169 bp in length
* 67649: gap of unknown length
* 67650 69178: contig of 1529 bp in length
* 69179 69198: gap of unknown length
* 69199 70095: contig of 897 bp in length
* 70096 70115: gap of unknown length
* 70116 71821: contig of 1706 bp in length
* 71822 71841: gap of unknown length
* 71842 73506: contig of 1665 bp in length
* 73507 73526: gap of unknown length
* 73527 74749: contig of 1223 bp in length
* 74750 74769: gap of unknown length
* 74770 75963: contig of 1194 bp in length
* 75964 75983: gap of unknown length
* 75984 77905: contig of 1922 bp in length
* 77906 77925: gap of unknown length
* 77926 79236: contig of 1311 bp in length
* 79237 79256: gap of unknown length
* 79257 81332: contig of 2076 bp in length
* 81333 81352: gap of unknown length
* 81353 83169: contig of 1817 bp in length
* 83170 83189: gap of unknown length
* 83190 84096: contig of 907 bp in length
* 84097 84116: gap of unknown length
* 84117 85749: contig of 1633 bp in length
* 85750 85769: gap of unknown length
* 85770 86770: contig of 1001 bp in length
* 86771 86790: gap of unknown length
* 86791 88207: contig of 1417 bp in length
* 88208 88227: gap of unknown length
* 88228 88976: contig of 749 bp in length
* 88977 88996: gap of unknown length
* 88997 89855: contig of 859 bp in length.
Location/Qualifiers
1..89855
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-76I16"
/sex="male"
1..7962
/note="assembly_name:Contig145"
7983..13924
/note="assembly_name:Contig144"
13945..18990
/note="assembly_name:Contig143"
19011..22501
/note="assembly_name:Contig142"
22522..27077
/note="assembly_name:Contig141"
27098..28819
/note="assembly_name:Contig140"
28840..31224
/note="assembly_name:Contig139"
31245..35093
/note="assembly_name:Contig138"
35114..38611
/note="assembly_name:Contig137"
38632..42261
/note="assembly_name:Contig136"
42282..44130
/note="assembly_name:Contig135"
44151..47059
/note="assembly_name:Contig134"
47080..49496
/note="assembly_name:Contig133"
49517..52236
/note="assembly_name:Contig132"
52257..54270
/note="assembly_name:Contig131"
54291..56271
/note="assembly_name:Contig130"

Fri May 24 10:49:46 2002

```
Db 140 TAAAAAGATAG-----TCCTCCCATTTAGCTAGAGAGCAACACCC-AGAAAGCCTATTGG 192
Qy 257 ttngtgcgcgcggccttgctggttgcgaagccgctagagctaccggcgagggg 316
Db 193 CTGCGCCCTCGCGGCGCTTGCTCGCTTTCAAGCGGCGCTCGGCTCGAGAGAGAGGC 252
Qy 317 cgggcgcgagctcgcgttgcgttggttaccagagacacacgtgcgcagctccggaacgg 376
Db 253 GGGCGGAGGCTAGCTGTGTGCTGCTCGGAGGACGCTGTCAGTCCCGGAAGCGG 312
Qy 377 ccgggggaagctgcctccgcgcgcctgcgcgagaaagcgcgcgggtccgctctgctc 436
Db 313 CGAGGGGAACCTGCTCGCGCGCGCGCGGAGGAGGAGCAACCGCGTCCCTTAGGGTCC 372
Qy 437 ggggtcggctggccatggatgccatgctgtagc 470
Db 373 GGGCGCGCGCGCCATGATCAATGCTGAGC 406

RESULT 3
LOCUS HS475N16/c
DEFINITION Human DNA sequence from clone RP3-475N16 on chromosome 6p12.3-21.2.
Contains the genes for CTG4A, pre-T cell receptor alpha, a novel
protein similar to RPL7A (60S ribosomal protein L7A) and the 3' end
of gene KIAA0240. Contains ESTs, STSs, GSSs and four putative CpG
islands, complete sequence.
ACCESSION AL035587
VERSION AL035587.5 GI:6002306
KEYWORDS HTG: CpG island; CTG4A; KIAA0240; RPL7A.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 113109)
Direct Submission
Submitted (02-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Request: clonerequest@sanger.ac.uk
On Sep 30, 1999 this sequence version replaced gi:5921383.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP3-475N16 is from the library RPCI-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-475N16.
Location/Qualifiers
1..113109
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p12.3-21.2"

FEATURES
source
/clone="RP3-475N16"
/clone_lib="RPCI-3"
1..70
/note="Alu repeat: matches 1..70 of consensus"
71..167
/note="MER51B repeat: matches 504..617 of consensus"
3540..3620
/note="L1MA4A repeat: matches 6197..6277 of consensus"
4703..4772
/note="Tigger2a repeat: matches 369..444 of consensus"
5085..5180
/note="Tigger2a repeat: matches 299..369 of consensus"
5476..5525
/note="Tigger2a repeat: matches 246..299 of consensus"
5844..6032
/note="Tigger2a repeat: matches 4..216 of consensus"
6563..6699
/note="MER91B repeat: matches 1..123 of consensus"
7000..7061
/note="MER91B repeat: matches 91..157 of consensus"
8465..8599
/note="FLAM_C repeat: matches 1..133 of consensus"
8832..8949
/note="FLAM_C repeat: matches 2..133 of consensus"
10905..11215
/note="AluSgl repeat: matches 1..309 of consensus"
complement(12117..22204)
/gene="dj475N16.1"
complement(12117)
/gene="dj475N16.1"
complement(join(12117..12836,13197..13314,13565..13687,
15734..15830,16806..16929,21683..22204))
/gene="dj475N16.1"
/note="match: cDNAs: Em:U80744 Em:X66602; match: ESTs:
Em:W27536 Em:AI878843 Em:D63172 Em:W52850 Em:AA639468
Em:AA002532 Em:AI413153 Em:AA048794 Em:AI992170
Em:AF150177 Em:AI391065 Em:AI464828 Em:N31024 Em:W29254
Em:AA772406 Em:AA468594 Em:AA292739 Em:AI050976
Em:AA261570 Em:W69640 Em:T69638 Em:Z97785 Em:C18340
Em:AI249023 Em:C00258 Em:AA419155 Em:AA315000 Em:AW007167
Em:AI215611 Em:AA065255 Em:T85390 Em:AA065254 Em:AA303388
Em:AI479630 Em:AI355887 Em:AA065252 Em:AI202110 Em:R24856
Em:AA065251 Em:AA456338 Em:AI193019 Em:W67832 Em:AI141843
Em:AA478239 Em:AI435295 Em:T12239 Em:AA810253 Em:AA251526
Em:AA527588 Em:AI762318 Em:AA446743 Em:T10814 Em:AA410700
Em:H44663 Em:H44749 Em:AA331641 Em:R44560 Em:AI356413
Em:155954 Em:AA065262 Em:AI386455 Em:AI887652 Em:AA065263
Em:AI143093 Em:AA378222 Em:AI093040 Em:AI095209"
/product="dj475N16.1 (CTG4A)"
/evidence=not_experimental
complement(12134)
/gene="dj475N16.1"
complement(12141)
/gene="dj475N16.1"
complement(join(12613..12836,13197..13314,13565..13687,
15734..15830,16806..16929,21683..21833))
/gene="dj475N16.1"
/note="owing to a frameshift at 15781 compared to cDNA
em:U80744 our translation differs considerably from the
published tr:O15412 protein, adding two more coding exons
to 3' or C-terminal end; match: proteins: Tr:O15412"
/codon_start=1
/evidence=not_experimental
/product="dj475N16.1 (CTG4A)"
/db_xref="CAB75301.1"
/db_xref="GI:6969163"
/translation="MDSMPASRCLLLPLLLLLLLPAPLGPQAGAEENDWVR
LPSCKECKYVAVELKSNFAPSTGKTKEVIGYGLDQKASGVKTKSDLRLIEVTET
ICKRLDLSLHKRGTSGNRKAGMSEFETLHNVHKGVKVMDIPYELWNETSAEVA
DLKQGDVLVEFEFVIEDWYRNHOEDLFLCNHVLKKGDTSCLAEOFGKGGDT
AALGKKKKKSSRAKAAAGRSSSSKQRKEELGGLGDEGSPEDDEGIQKASPLTHSPPD
EL"
15046..15456
misc_feature
```

misc_feature	/note="match: STS: Em:G37319" complement(17091..17133) /gene="dj475N16.1" /note="match: STS: Em:G25145 Em:G27441" 21516..22025 /note="CpG island" /evidence=not_experimental 21749..21784 /note="match: STS: Em:AF021121" complement(21753..21788) /gene="dj475N16.1" /note="match: STS: Em:G09564" 21759..21794 /note="match: STS: Em:267474" 21762..21797 /note="match: STS: Em:G09799" complement(21762..21793) /gene="dj475N16.1" /note="match: STS: Em:G09544" complement(22138..22473) /note="match: GSS: Em:AQ756680" 23412..23718 /note="AluSc repeat: matches 2..309 of consensus" complement(23719..23975) /note="match: GSS: Em:AQ267139" complement(24876..25553) /note="match: GSS: Em:AQ383009" complement(25571) /gene="dj475N16.2" complement(join(25571..26188,27131..27175,28057..28377,35277..35438)) /note="match: cDNAs: Em:U16958 Em:U36759 Em:U38996; match: STS: Em:AA913164 Em:AI263200 Em:AI285999 Em:AW014894 Em:AI767176" /product="dj475N16.2 (pre-T cell receptor alpha)" /evidence=not_experimental complement(25571..35438) /gene="dj475N16.2" complement(25590..25595) /gene="dj475N16.2" complement(join(25722..26188,27131..27175,28057..28377,35277..35334)) /gene="dj475N16.2" /codon_start=1 /evidence=not_experimental /product="dj475N16.2 (pre-T cell receptor alpha)" /protein_id="CAB75302.1" /db_xref="GI:6969164" /translation="MAGTWLLLLALGALPTGVGGTFFSLAPPTMLLVGQKQWV EPLVCHTGPAGHSRSTPWFSGNSALDAFTYGPSATDGTWNLHLSLSESLASW GALNLGVLRLFLLLFDLLTSCSLDPAFLPSFATTLRLALGSHRLHPATETG GKEATSPRPQPRRRWGDTTPPKRPGSPVWGEYSYSSYPTCPAQCWSSRLRAPSS SLGAFAGDLPPLQAGAA" 29689..29994 /note="AluSc repeat: matches 1..306 of consensus" 30593..30668 /note="L1MB8 repeat: matches 6101..6170 of consensus" 33433..33494 /note="31 copies 2 mer tt 69% conserved" 38170..38299 /note="L1MB3 repeat: matches 6024..6137 of consensus" 38610..38858 /note="L1MB3 repeat: matches 5711..6024 of consensus" 38941..39002 /note="AluU/FAM repeat: matches 19..80 of consensus" 39289..39862 /note="CpG island" /evidence=not_experimental 40745..40851 /note="L1B4 repeat: matches 5703..5817 of consensus" 41336..41377
misc_feature	/note="match: STS: Em:G37319" complement(17091..17133) /gene="dj475N16.1" /note="match: STS: Em:G25145 Em:G27441" 21516..22025 /note="CpG island" /evidence=not_experimental 21749..21784 /note="match: STS: Em:AF021121" complement(21753..21788) /gene="dj475N16.1" /note="match: STS: Em:G09564" 21759..21794 /note="match: STS: Em:267474" 21762..21797 /note="match: STS: Em:G09799" complement(21762..21793) /gene="dj475N16.1" /note="match: STS: Em:G09544" complement(22138..22473) /note="match: GSS: Em:AQ756680" 23412..23718 /note="AluSc repeat: matches 2..309 of consensus" complement(23719..23975) /note="match: GSS: Em:AQ267139" complement(24876..25553) /note="match: GSS: Em:AQ383009" complement(25571) /gene="dj475N16.2" complement(join(25571..26188,27131..27175,28057..28377,35277..35438)) /note="match: cDNAs: Em:U16958 Em:U36759 Em:U38996; match: STS: Em:AA913164 Em:AI263200 Em:AI285999 Em:AW014894 Em:AI767176" /product="dj475N16.2 (pre-T cell receptor alpha)" /evidence=not_experimental complement(25571..35438) /gene="dj475N16.2" complement(25590..25595) /gene="dj475N16.2" complement(join(25722..26188,27131..27175,28057..28377,35277..35334)) /gene="dj475N16.2" /codon_start=1 /evidence=not_experimental /product="dj475N16.2 (pre-T cell receptor alpha)" /protein_id="CAB75302.1" /db_xref="GI:6969164" /translation="MAGTWLLLLALGALPTGVGGTFFSLAPPTMLLVGQKQWV EPLVCHTGPAGHSRSTPWFSGNSALDAFTYGPSATDGTWNLHLSLSESLASW GALNLGVLRLFLLLFDLLTSCSLDPAFLPSFATTLRLALGSHRLHPATETG GKEATSPRPQPRRRWGDTTPPKRPGSPVWGEYSYSSYPTCPAQCWSSRLRAPSS SLGAFAGDLPPLQAGAA" 29689..29994 /note="AluSc repeat: matches 1..306 of consensus" 30593..30668 /note="L1MB8 repeat: matches 6101..6170 of consensus" 33433..33494 /note="31 copies 2 mer tt 69% conserved" 38170..38299 /note="L1MB3 repeat: matches 6024..6137 of consensus" 38610..38858 /note="L1MB3 repeat: matches 5711..6024 of consensus" 38941..39002 /note="AluU/FAM repeat: matches 19..80 of consensus" 39289..39862 /note="CpG island" /evidence=not_experimental 40745..40851 /note="L1B4 repeat: matches 5703..5817 of consensus" 41336..41377
misc_feature	/note="match: STS: Em:G37319" complement(17091..17133) /gene="dj475N16.1" /note="match: STS: Em:G25145 Em:G27441" 21516..22025 /note="CpG island" /evidence=not_experimental 21749..21784 /note="match: STS: Em:AF021121" complement(21753..21788) /gene="dj475N16.1" /note="match: STS: Em:G09564" 21759..21794 /note="match: STS: Em:267474" 21762..21797 /note="match: STS: Em:G09799" complement(21762..21793) /gene="dj475N16.1" /note="match: STS: Em:G09544" complement(22138..22473) /note="match: GSS: Em:AQ756680" 23412..23718 /note="AluSc repeat: matches 2..309 of consensus" complement(23719..23975) /note="match: GSS: Em:AQ267139" complement(24876..25553) /note="match: GSS: Em:AQ383009" complement(25571) /gene="dj475N16.2" complement(join(25571..26188,27131..27175,28057..28377,35277..35438)) /note="match: cDNAs: Em:U16958 Em:U36759 Em:U38996; match: STS: Em:AA913164 Em:AI263200 Em:AI285999 Em:AW014894 Em:AI767176" /product="dj475N16.2 (pre-T cell receptor alpha)" /evidence=not_experimental complement(25571..35438) /gene="dj475N16.2" complement(25590..25595) /gene="dj475N16.2" complement(join(25722..26188,27131..27175,28057..28377,35277..35334)) /gene="dj475N16.2" /codon_start=1 /evidence=not_experimental /product="dj475N16.2 (pre-T cell receptor alpha)" /protein_id="CAB75302.1" /db_xref="GI:6969164" /translation="MAGTWLLLLALGALPTGVGGTFFSLAPPTMLLVGQKQWV EPLVCHTGPAGHSRSTPWFSGNSALDAFTYGPSATDGTWNLHLSLSESLASW GALNLGVLRLFLLLFDLLTSCSLDPAFLPSFATTLRLALGSHRLHPATETG GKEATSPRPQPRRRWGDTTPPKRPGSPVWGEYSYSSYPTCPAQCWSSRLRAPSS SLGAFAGDLPPLQAGAA" 29689..29994 /note="AluSc repeat: matches 1..306 of consensus" 30593..30668 /note="L1MB8 repeat: matches 6101..6170 of consensus" 33433..33494 /note="31 copies 2 mer tt 69% conserved" 38170..38299 /note="L1MB3 repeat: matches 6024..6137 of consensus" 38610..38858 /note="L1MB3 repeat: matches 5711..6024 of consensus" 38941..39002 /note="AluU/FAM repeat: matches 19..80 of consensus" 39289..39862 /note="CpG island" /evidence=not_experimental 40745..40851 /note="L1B4 repeat: matches 5703..5817 of consensus" 41336..41377
misc_feature	/note="match: STS: Em:G37319" complement(17091..17133) /gene="dj475N16.1" /note="match: STS: Em:G25145 Em:G27441" 21516..22025 /note="CpG island" /evidence=not_experimental 21749..21784 /note="match: STS: Em:AF021121" complement(21753..21788) /gene="dj475N16.1" /note="match: STS: Em:G09564" 21759..21794 /note="match: STS: Em:267474" 21762..21797 /note="match: STS: Em:G09799" complement(21762..21793) /gene="dj475N16.1" /note="match: STS: Em:G09544" complement(22138..22473) /note="match: GSS: Em:AQ756680" 23412..23718 /note="AluSc repeat: matches 2..309 of consensus" complement(23719..23975) /note="match: GSS: Em:AQ267139" complement(24876..25553) /note="match: GSS: Em:AQ383009" complement(25571) /gene="dj475N16.2" complement(join(25571..26188,27131..27175,28057..28377,35277..35438)) /note="match: cDNAs: Em:U16958 Em:U36759 Em:U38996; match: STS: Em:AA913164 Em:AI263200 Em:AI285999 Em:AW014894 Em:AI767176" /product="

FEATURES
source

BASE COUNT 351 a 400 c 422 g 249 t

ORIGIN

Query Match 13.2%; Score 62.4; DB 9; Length 1422;

Best Local Similarity 78.1%; Pred. No. 0.00026;

Matches 75; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 375 gggccgggaagctgctcgccgctgcggaggaagcgcggcggtgcctctgtct 434

DB 1 ggcacaggaactgctcgccgctgcggaggaagcgcggcggtgcctctgtct 60

OY 435 ctgggtccgctggcgcagtgatctgctgagc 470

DB 61 cggcgccggcgccgagcagtgatctgagc 96

RESULT 9

AC090906/c

LOCUS

DEFINITION Homo sapiens chromosome 15 clone RP11-430B1 map 15, LOW-PASS

SEQUENCE SAMPLING.

AC090906

AC090906.1 GI:13374684

VERSION

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 60424)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 15, clone RP11-430B1

Unpublished

2 (bases 1 to 60424)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zemlek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smir, A. F. A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L13133

Center clone name: 430_B_1

NOTE: This record contains 76 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

1 705: contig of 705 bp in length

706 805: gap of 100 bp

806 1508: contig of 703 bp in length

1509 1608: gap of 100 bp

1609 2298: contig of 690 bp in length

2299 2398: gap of 100 bp

2399 3080: contig of 682 bp in length

3081 3180: gap of 100 bp

3181 3877: contig of 697 bp in length

3878 3977: gap of 100 bp

3978 4668: contig of 691 bp in length

4669 4768: gap of 100 bp

4769 5472: contig of 704 bp in length

5473 5572: gap of 100 bp

5573 6279: contig of 707 bp in length

6280 6379: gap of 100 bp

6380 7082: contig of 703 bp in length

7083 7182: gap of 100 bp

7183 7870: contig of 688 bp in length

7871 7970: gap of 100 bp

7971 8672: contig of 702 bp in length

8673 8772: gap of 100 bp

8773 9455: contig of 683 bp in length

9456 9555: gap of 100 bp

9556 10257: contig of 702 bp in length

10258 10357: gap of 100 bp

10358 11051: contig of 694 bp in length

11052 11151: gap of 100 bp

11152 11813: contig of 662 bp in length

11814 11913: gap of 100 bp

11914 12619: contig of 706 bp in length

12620 12719: gap of 100 bp

12720 13413: contig of 694 bp in length

13414 13513: gap of 100 bp

13514 14222: contig of 709 bp in length

14223 14322: gap of 100 bp

14323 15018: contig of 696 bp in length

15019 15118: gap of 100 bp

15119 15808: contig of 690 bp in length

15809 15908: gap of 100 bp

15909 16610: contig of 702 bp in length

16611 16710: gap of 100 bp

16711 17416: contig of 706 bp in length

17417 17516: gap of 100 bp

17517 18203: contig of 687 bp in length

18204 18303: gap of 100 bp

18304 19006: contig of 703 bp in length

19007 19106: gap of 100 bp

19107 19799: contig of 693 bp in length

19800 19899: gap of 100 bp

19900 20564: contig of 665 bp in length

20565 20664: gap of 100 bp

20665 21353: contig of 689 bp in length

21354 21453: gap of 100 bp

21454 22158: contig of 705 bp in length

22159 22258: gap of 100 bp

22259 22950: contig of 692 bp in length

22951 23050: gap of 100 bp

23051 23746: contig of 696 bp in length

23747 23846: gap of 100 bp

gene

Web site: <http://www.jgi.doe.gov>

 Project Information
 Center Project Name: 0
 Center Clone Name: RPCI-23_41F9

Summary Statistics

Consensus quality: 169456 bases at least Q40
 Consensus quality: 196444 bases at least Q30
 Consensus quality: 205812 bases at least Q20
 Estimated insert size: 226654; sum-of-contigs estimation
 Quality coverage: 3.43 in Q20 bases; agarose-fp estimation
 Quality coverage: 3.19 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 37 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 1422: contig of 1422 bp in length
* 1423 1522: gap of unknown length
* 1523 3563: contig of 2041 bp in length
* 3564 3663: gap of unknown length
* 3664 4879: contig of 1216 bp in length
* 4880 4979: gap of unknown length
* 4980 6061: contig of 1082 bp in length
* 6062 6161: gap of unknown length
* 6162 7308: contig of 1147 bp in length
* 7309 7409: gap of unknown length
* 7409 8607: contig of 1199 bp in length
* 8608 8707: gap of unknown length
* 8708 10851: contig of 2144 bp in length
* 10852 12535: contig of 1584 bp in length
* 12536 12635: gap of unknown length
* 12636 13837: contig of 1202 bp in length
* 13838 13937: gap of unknown length
* 13938 16329: contig of 2392 bp in length
* 16330 16429: gap of unknown length
* 16430 18243: contig of 1813 bp in length
* 18243 21090: gap of unknown length
* 21090 21190: contig of 2748 bp in length
* 21191 23891: contig of 2701 bp in length
* 23892 23991: gap of unknown length
* 23992 26366: contig of 2375 bp in length
* 26367 29744: contig of 3278 bp in length
* 29745 29844: gap of unknown length
* 29845 33746: contig of 3902 bp in length
* 33747 33846: gap of unknown length
* 33847 38047: contig of 4201 bp in length
* 38048 38147: gap of unknown length
* 38148 41062: contig of 2915 bp in length
* 41063 41162: gap of unknown length
* 41163 44603: contig of 3441 bp in length
* 44604 44703: gap of unknown length
* 44704 47610: contig of 2907 bp in length
* 47611 47710: gap of unknown length
* 47711 51750: contig of 4040 bp in length
* 51751 51850: gap of unknown length
* 51851 56139: contig of 4289 bp in length
* 56140 56239: gap of unknown length
* 56240 63693: contig of 7454 bp in length
* 63694 63793: gap of unknown length
* 63794 68500: contig of 4707 bp in length
* 68501 68600: gap of unknown length
* 68601 75767: contig of 7167 bp in length
* 75768 82947: contig of 7080 bp in length

```

```

* 82948 83048 83047: gap of unknown length
* 83048 89217: contig of 6170 bp in length
* 89218 93117: gap of unknown length
* 93118 96269: contig of 6952 bp in length
* 96270 96369: gap of unknown length
* 96370 102865: contig of 6496 bp in length
* 102866 102966: gap of unknown length
* 102967 113573: contig of 10608 bp in length
* 113574 121561: gap of unknown length
* 121562 121661: contig of 7888 bp in length
* 121662 131831: contig of 10170 bp in length
* 131832 131931: gap of unknown length
* 131932 143351: contig of 11420 bp in length
* 143352 143451: gap of unknown length
* 143452 154592: contig of 11141 bp in length
* 154593 154692: gap of unknown length
* 154693 165992: contig of 11300 bp in length
* 165993 166092: gap of unknown length
* 166093 189174: contig of 23082 bp in length
* 189175 189274: gap of unknown length
* 189275 230254: contig of 40980 bp in length.

```

FEATURES

```

Location/Qualifiers
1..230254
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone="RP23-41F9"
  /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 61269 a 52644 c 52155 g 60320 t 3866 others
ORIGIN

```

```

Query Match          9.4%; Score 44.4; DB 2; Length 230254;
Best Local Similarity 54.2%; Pred. No. 4.1;
Matches 90; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```

```

QY 268 gcgggcttggtcggttcgaagccgctagaggtacccggcgagggcgagggcgagc 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40774 GGGGGTGGCGCGTGGCGCGGGGGCGGCGCTCCCGCGCGGTGGGGGGCTCGCGG 40833
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 328 tcgcgtgctggtttaccagagacacgtgcgagtcctccggaagcggcgagggaagc 387
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40834 CGCGCGGGCGGTGGGTGGCGCGCCGACACGGGGCGGCGCGGTCCGCCCGCTTGG 40893
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 388 tcgtcgcgcgcctcgcggaggaagcgcgcgcgggtccgcctcgc 433
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40894 GGCTCCCGCGGCTCGGCGCGGTGGGGTGGCGCGCCCGCGCGC 40939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 13

```

SCD78 Streptomyces coelicolor cosmid D78. linear BCT 26-NOV-1998
LOCUS Streptomyces coelicolor cosmid D78.
DEFINITION AL034355
ACCESSION AL034355.1 GI:3928709
VERSION
KEYWORDS
ABC transporter; ftsK/spoIIIE family protein; hsc2;
histidinol-phosphate aminotransferase; integrase/recombinase;
integrated element; oxidoreductase; partial thiamin biosynthesis
protein thic; pheA; plasmid replication initiator protein;
prephenate dehydratase; secreted peptidase; serine/threonine
phosphatase; serS; seryl-tRNA synthase; thic'; two component
sensor.

```

SOURCE

```

ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 36224)
Redenbach, M., Kieser, H.M., Denapalte, D., Eichner, A., Cullum, J.,
Kinashi, H. and Hopwood, D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351

```

REFERENCE
Saunders,D.C. and Harris,D.
AUTHORS
Unpublished
REFERENCE
Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE
Submitted (25-NOV-1998) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrel@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/projects/S_coelicolor/) CDS are
numbered using the following system eg SC787.01c. SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.gov/jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid D78 lies to
the right of Q11 on the Asel-D genomic restriction fragment.

FEATURES
source
1..36224
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid D78"
1..817
/gene="SCD78.01"
<1..817
/note="SCD78.01"
/note="SCD78.01, ftsK /spoIIIE family protein, partial CDS
within putative integrated plasmid, len: 271 aa; similar
to many members of the ftsK /spoIIIE family e.g.SPI_STRAZ
(EMBL:S58719) sporulation-regulatory protein from
Streptomyces azureus plasmid pSal.1 (303 aa) fasta
scores; opt: 736 z-score: 1014.5 E(): 0, 49.0% identity in
247 aa overlap and TRSA_STRAM (EMBL:Z19593) transfer
protein traSA from Streptomyces ambifaciens plasmid pSAM2
(306 aa), fasta scores; opt: 222 z-score: 676.5 E():
2e-30, 47.8% identity in 203 aa overlap. Also similar to
SC6A9.34, a protein within another putative integrated
plasmid (on Streptomyces coelicolor cosmid St6A)(E):
5.4e-24, 45.7% identity in 247 aa overlap). Contains
PS00017 ATP /GTP-binding site motif A (P-loop)
/codon_start=2
/transl_table=11
/product="ftsK /spoIIIE family protein"
/protein_id="CAA22206.1"
/db_xref="GI:3928710"
/translation="IGALETGGAWNNRLVPWHLIAGATRSKGSTLLARVTITQAPQ

misc_feature
1..4443
/note="possible integrated element"

misc_feature
71..94
/gene="SCD78.01"
/note="PS00017 ATP /GTP-binding site motif A (P-loop)"

gene
814..1011
/gene="SCD78.02"
814..1011
/note="SCD78.02"
/note="SCD78.02, small hydrophobic protein, len: 65 aa;
within putative integrated plasmid"

CDS
/codon_start=1
/transl_table=11
/product="hypothetical protein SCD78.02"
/protein_id="CAA22207.1"
/db_xref="GI:3928711"
/translation="MISEGTAFGFVVGFIIVLLVSRDRVRGWAEVCVGLFLGYLCLO
TPELFTIHGLTVTWISGSFHT"

gene
1308..2666
/gene="SCD78.03"
1308..2666
/partial

CDS
/gene="SCD78.03"
/note="SCD78.03, possible plasmid replication initiator
protein within putative integrated plasmid, len: 452 aa;
similar to REPS_STRAM (EMBL:Z19594) replication initiator
protein from Streptomyces ambifaciens plasmid pSAM2 (459
aa), fasta scores; opt: 712 z-score: 983.7 E(): 0, (43.7%
identity in 467 aa overlap). Note that there is no
possible initiation codon upstream of the start of
homology (or close downstream). The CDS given here has an
arbitrary start"

misc_feature
/codon_start=1
/transl_table=11
/product="putative plasmid replication initiator protein"
/protein_id="CAA22208.1"
/db_xref="GI:3928712"
/translation="MRHLPEDRDLRIADDPFRWLQOITATGCCAHPLVLSGHTT
TLDNATGETILHYDTNEPGRLLVCRNRSTVCGPCSRLEHADPTFLHVLAGLLGCK
NVPASVTRPRLFVLITAPSGFAVHRAGOTCRPRDGCGAGHPGLCGCTVHAIDPADPA
VGOLPCDCYDYTAHVLMHAHASKLWDREVIDRRRLASSAGI VOSREFAHHARLSFA
IAYOKRAAVHHVTHVRLVDGPDPGTDEPPANGVTDRLTRAVHASAQRTVHTPTSLAT
GELELSKGTQTDVCLHGKDGDADDAAVAAYVAKYVSKGTDTAGTGDKHTVTTCCDDIID
SARVSHVTLMTCHRWLGCLPEYAPLRLTWTHTLTGVRGHILT KSRAYSITTYALRSY
QRAHHGHDTDTDAITDAHWRVVGSGHFGAALIAGAEDIAQNQAARAARDALSAGGS
AT"

gene
complement(2805..4292)
/gene="SCD78.04c"
complement(2805..4292)

CDS
/gene="SCD78.04c"
/note="SCD78.04c, possible integrase/ recombinase within
putative integrated plasmid, len: 495 aa; similar to
TR:Q37839 (EMBL:D38173) ORF469 protein (possible
site-specific recombinase) from Actinophaga R4 (469 aa);
fasta scores; opt: 261 z-score: 320.1 E(): 1.4e-10, 29.1%
identity in 306 aa overlap and TR:O06604 (EMBL:Z95586)
MYCY336.18 (Rv1586c) possible integrase from M.
tuberculosis integrated phage phiRv1 (469 aa), fasta
scores; opt: 371 z-score: 270.8 E(): 7.8e-08, (26.6%
identity in 489 aa overlap). Probable coiled-coil from 41
to 451 (38 residues) Max score: 1.608 (probability 0.99%)
/codon_start=1
/transl_table=11
/product="putative integrase /recombinase"
/protein_id="CAA22209.1"
/db_xref="GI:3928713"
/translation="MLAADPNLKVCVYARI SPDGRVKDHGIEDQHDMSEARRFSY
WLIVRYTNDKSKASESVVRDDFEQLADLAAGATPEGYPVHWNVMDRLRRRP
DWERYLKAFTSQDGRVYHDSNGLDLYAEGFEIKGLGVGVMASLSETRKKQRKRNSRH

REFERENCE	AUTHORS	JOURNAL	TITLE	COMMENT
1	(bases 1 to 205914)	DOE Joint Genome Institute.	Sequencing of Human Chromosome 16	
2	(bases 1 to 205914)	DOE Joint Genome Institute.	Direct Submission	
		Submitted (01-APR-2000)	Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
		On April 25, 2001 this sequence version replaced gi:10281392.		
		-----Genome Center		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 21:53:03 ; Search time 525.6 Seconds
(without alignments)
81.664 Million cell updates/sec

Title: US-09-696-686-45

Perfect score: 25

Sequence: 1 agctcagacatggactccatggccc 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	22	AAS05280
2	25	100.0	25	22	AAS05287
3	25	100.0	49	22	AAS05282
4	25	100.0	471	22	AAS05284
5	25	100.0	1839	22	AAS05278
6	25	100.0	1848	22	AAS05286
7	19.2	76.8	1675	21	AAS21311
8	18.6	74.4	435	21	AA27030
9	18.6	74.4	585	22	AAS26338

C 10	18.6	74.4	933	22	AAS26339	Human cDNA encodin
C 11	18.6	74.4	1422	21	AAA27028	Human cbl-SL cDNA
C 12	18.6	74.4	1522	22	AAS25883	Human cDNA encodin
C 13	18.6	74.4	1526	22	AAK51619	Human polynucleoti
C 14	18.6	74.4	1543	22	AAK52603	Human polynucleoti
C 15	18.6	74.4	1547	21	AAK27027	Human cbl-SL cDNA.
C 16	18.6	74.4	2964	23	AAS89487	DNA encoding novel
C 17	18.2	72.8	1291	21	AAC49165	Arabidopsis thalia
C 18	18.2	72.8	1295	21	AAC35699	Arabidopsis thalia
C 19	18.2	72.8	30598	23	ABL01908	Drosophila melanog
C 20	17.8	71.2	2384	22	ABD21727	Mouse stem cell gr
C 21	17.8	71.2	54548	21	AAZ45596	DNA sequence of th
C 22	17.6	70.4	157	22	ABA74959	Human foetal liver
C 23	17.6	70.4	157	22	ABA39649	Human foetal liver
C 24	17.6	70.4	157	22	AAK23460	Human brain expres
C 25	17.6	70.4	157	22	AAK49606	Human bone marrow
C 26	17.6	70.4	157	22	AAI26712	Probe #16645 for g
C 27	17.6	70.4	157	22	AAI55487	Probe #24173 used
C 28	17.6	70.4	583	22	ABA62434	Human foetal liver
C 29	17.6	70.4	583	22	ABA29768	Probe #8234 for ge
C 30	17.6	70.4	583	22	AAK10776	Human brain expres
C 31	17.6	70.4	583	22	AAK36652	Human bone marrow
C 32	17.6	70.4	583	22	AAI17503	Probe #7436 for ge
C 33	17.6	70.4	583	22	AAI42414	Probe #11100 used
C 34	17.6	70.4	873	22	AAH46902	CDNA encoding huma
C 35	17.6	70.4	1118	23	AAS85770	DNA encoding novel
C 36	17.6	70.4	1292	14	AAQ47726	Coding sequence of
C 37	17.6	70.4	1327	24	AAS94754	Human DNA sequence
C 38	17.6	70.4	1362	22	AAS05279	Human trinucleotid
C 39	17.6	70.4	1462	20	AAZ24833	Human secreted pro
C 40	17.6	70.4	1478	24	AB199729	Mouse ischaemic co
C 41	17.6	70.4	1515	21	AAA87711	Secreted protein e
C 42	17.6	70.4	1515	21	AAA87748	Human secreted pro
C 43	17.6	70.4	1515	22	AAF63993	CDNA encoding secr
C 44	17.6	70.4	1515	22	AAF64030	CDNA encoding huma
C 45	17.6	70.4	1532	21	AAZ64995	Membrane-bound pro

ALIGNMENTS

RESULT 1

AAS05280

ID AAS05280 standard; DNA; 25 BP.

XX AAS05280;

AC AAS05280;

XX 07-SEP-2001 (first entry)

DT PCR primer #1 used to amplify T243 homologous sequences.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;

DE transgenic animal; knockout mouse; triplet repeat expansion;

KW fragile X syndrome; Huntington's disease; mouse; PCR primer; sa.

XX Mus sp.

OS WO200130798-A1.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-US29382.

XX 26-OCT-1999; 99US-0161488.

PR (DELT-) DELTAGEN INC.

PA Klein R, Matthews W, Moore M, Allen KD;

PI WPI; 2001-300473/31.

XX Novel transgenic animals useful as animal model for characterization of

PT function of a gene encoding trinucleotide repeat proteins (TRPs),

PT contains heterozygous disruption in a gene encoding TRP -
 XX Claim 13; Fig 13; 106pp; English.
 PS The present sequence for PCR primer #1 is used to amplify sequences
 CC homologous to the T243 gene which encodes for a trinucleotide repeat
 CC protein (TRP). The invention describes methods of producing embryonic
 CC stem (ES) cells comprising a heterozygous disruption in a target DNA
 CC sequence (preferably gene T243) encoding a TRP and of producing a
 CC knockout mouse comprising a homozygous disruption in a gene encoding TRP,
 CC where the disruption inhibits the production of the wild type TRP. The
 CC invention also relates to identifying agents capable of affecting a
 CC phenotype of a knockout mouse. Also described are methods of determining
 CC whether expansion of the trinucleotide repeat in a gene encoding TRP
 CC produces a phenotypic change. The transgenic animals and the cells are
 CC useful for identifying compounds capable of ameliorating disease
 CC symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.
 XX
 SQ Sequence 25 BP; 6 A; 9 C; 6 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.066;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 agctcagacatgagctccatgccc 25
 Db 1 agctcagacatgagctccatgccc 25

RESULT 2
 AAS05287/C
 ID AAS05287 standard; DNA; 25 BP.
 XX
 AC AAS05287;
 XX
 DT 07-SEP-2001 (first entry)
 DE Mouse ES cell T243 gene PCR primer #426.
 XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW mouse; PCR primer; ss.
 XX Mus sp.
 OS WO200130798-A1.
 XX
 PN 03-MAY-2001.
 XX
 PD 26-OCT-2000; 2000WO-US29382.
 XX
 PF 26-OCT-1999; 99US-0161488.
 XX
 PR (DELT-) DELTAGEN INC.
 XX
 PA Klein R, Matthews W, Moore M, Allen KD;
 XX WPI; 2001-300473/31.
 XX Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 PT contains heterozygous disruption in a gene encoding TRP -
 XX Example 12; Fig 13; 106pp; English.
 PS

XX The present sequence for PCR primer #426 is used to amplify the T243
 CC gene from an R1 mouse embryonic stem cell genomic library. The T243 gene
 CC encodes for a trinucleotide repeat protein (TRP). The invention describes
 CC methods of producing embryonic stem (ES) cells comprising a heterozygous
 CC disruption in a target DNA sequence (preferably T243) encoding a TRP and
 CC of producing a knockout mouse comprising a homozygous disruption in a
 CC gene encoding TRP, where the disruption inhibits the production of the
 CC wild type TRP. The invention also relates to identifying agents capable
 CC of affecting a phenotype of a knockout mouse. Also described are methods
 CC of determining whether expansion of the trinucleotide repeat in a gene
 CC encoding TRP produces a phenotypic change. The transgenic animals and
 CC the cells are useful for identifying compounds capable of ameliorating
 CC disease symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.
 XX
 SQ Sequence 25 BP; 4 A; 6 C; 9 G; 6 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.066;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 agctcagacatgagctccatgccc 25
 Db 25 agctcagacatgagctccatgccc 1

RESULT 3
 AAS05282
 ID AAS05282 standard; DNA; 49 BP.
 XX
 AC AAS05282;
 XX
 DT 07-SEP-2001 (first entry)
 DE PCR primer #488 used to amplify T243 PCR product from pool H10.
 XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; mouse; PCR primer; ss.
 XX Mus sp.
 OS WO200130798-A1.
 XX
 PN 03-MAY-2001.
 XX
 PD 26-OCT-2000; 2000WO-US29382.
 XX
 PF 26-OCT-1999; 99US-0161488.
 XX
 PR (DELT-) DELTAGEN INC.
 XX
 PA Klein R, Matthews W, Moore M, Allen KD;
 XX WPI; 2001-300473/31.
 XX Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 PT contains heterozygous disruption in a gene encoding TRP -
 XX Example 12; Fig 13; 106pp; English.
 PS The present sequence for PCR primer #488 is used to further amplify
 CC a T243 gene PCR product from pool H10 (obtained by PCR of the T243 gene
 CC

CC from an R1 mouse embryonic stem cell genomic library). Gene T243 encodes
 CC for a trinucleotide repeat protein (TRP). The invention describes methods
 CC of producing embryonic stem (ES) cells comprising a heterozygous
 CC disruption in a target DNA sequence (preferably gene T243) encoding a
 CC TRP and of producing a knockout mouse comprising a homozygous disruption
 CC in a gene encoding TRP, where the disruption inhibits the production of
 CC the wild type TRP. The invention also relates to identifying agents
 CC capable of affecting a phenotype of a knockout mouse. The transgenic
 CC animals and the cells are useful for identifying compounds capable of
 CC ameliorating disease symptoms, and as test substrates for the
 CC identification of drugs, pharmaceuticals, therapies and interventions
 CC which may be effective in treating trinucleotide repeat disorders e.g.
 CC fragile X syndrome and Huntington's disease. The animal models for
 CC trinucleotide repeat disorders are ideal model systems to study the
 CC progression of disease in vivo, the molecular basis of these diseases
 CC and show the features observed in human disease. Using the mice, it is
 CC possible to model both the pathogenic mechanism and the trinucleotide
 CC repeat instability in the mouse.

XX
 SQ Sequence 49 BP; 8 A; 16 C; 13 G; 12 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 49;
 Best Local Similarity 100.0%; Pred. No. 0.071;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 |||
 Db 25 agctcagacatggactccatggccc 49

RESULT 4

AAS05284/c
 ID AAS05284 standard; DNA; 471 BP.

XX AC AAS05284;

XX DT 07-SEP-2001 (first entry)

XX DE T243 gene homologous sequence #1 generated by PCR.

XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 XX transgenic animal; knockout mouse; triplet repeat expansion;
 XX fragile X syndrome; Huntington's disease; mouse; ds.

XX OS Mus sp.

XX PN WO200130798-A1.

XX PD 03-MAY-2001.

XX PF 26-OCT-2000; 2000WO-US29382.

XX PR 26-OCT-1999; 99US-0161488.

XX PA (DELT-) DELTAGEN INC.

XX PI Klein R, Matthews W, Moore M, Allen KD;

XX DR WPI; 2001-300473/31.

XX PT Novel transgenic animals useful as animal model for characterization of
 XX function of a gene encoding trinucleotide repeat proteins (TRPs),
 XX contains heterozygous disruption in a gene encoding TRP -

XX PS Claim 11; Fig 14; 106pp; English.

XX CC The present sequence for T243 homologous sequence #1 is generated by PCR
 XX and is homologous to the T243 gene which encodes for a trinucleotide
 XX repeat protein (TRP). The invention describes methods of producing
 XX embryonic stem (ES) cells comprising a heterozygous disruption in a
 XX target DNA sequence (preferably gene T243) encoding a TRP and of
 XX producing a knockout mouse comprising a homozygous disruption in a gene

CC encoding TRP, where the disruption inhibits the production of the wild
 CC type TRP. The invention also relates to identifying agents capable of
 CC affecting a phenotype of a knockout mouse. Also described are methods of
 CC determining whether expansion of the trinucleotide repeat in a gene
 CC encoding TRP produces a phenotypic change. The transgenic animals and the
 CC cells are useful for identifying compounds capable of ameliorating
 CC disease symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.

SQ Sequence 471 BP; 105 A; 129 C; 148 G; 88 T; 1 other;

Query Match 100.0%; Score 25; DB 22; Length 471;
 Best Local Similarity 100.0%; Pred. No. 0.092;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 |||
 Db 471 AGCTCAGACATGGACTCCATGGCCC 447

RESULT 5

AAS05278/c
 ID AAS05278 standard; cDNA; 1839 BP.

XX AC AAS05278;

XX DT 07-SEP-2001 (first entry)

XX DE Murine trinucleotide repeat protein (TRP) cDNA sequence.

XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 XX transgenic animal; knockout mouse; triplet repeat expansion;
 XX fragile X syndrome; Huntington's disease; mouse; murine; ss.

XX OS Mus sp.

XX FH Key Location/Qualifiers
 XX CDS 55..885

XX FT /*tag= a
 XX FT /product= "TRP"

XX PN WO200130798-A1.

XX PD 03-MAY-2001.

XX PF 26-OCT-2000; 2000WO-US29382.

XX PR 26-OCT-1999; 99US-0161488.

XX PA (DELT-) DELTAGEN INC.

XX PI Klein R, Matthews W, Moore M, Allen KD;

XX DR WPI; 2001-300473/31.

XX DR P-PSDB; AAU02498.

XX PT Novel transgenic animals useful as animal model for characterization of
 XX function of a gene encoding trinucleotide repeat proteins (TRPs),
 XX contains heterozygous disruption in a gene encoding TRP -
 XX Claim 8; Fig 11; 106pp; English.

XX CC The present sequence encodes for murine trinucleotide repeat protein
 XX (TRP). The invention describes methods of producing embryonic stem (ES)
 XX cells comprising a heterozygous disruption in a target DNA sequence

Fri May 24 10:49:36 2002

us-09-696-686-45.rng

CC (preferably gene T243) encoding a TRP and of producing a knockout mouse
 CC comprising a homozygous disruption in a gene encoding TRP, where the
 CC disruption inhibits the production of the wild type TRP. The invention
 CC also relates to identifying agents capable of affecting a phenotype of
 CC a knockout mouse. Also described are methods of determining whether
 CC expansion of the trinucleotide repeat in a gene encoding TRP produces a
 CC phenotypic change. The transgenic animals and the cells are useful for
 CC identifying compounds capable of ameliorating disease symptoms, and as
 CC test substrates for the identification of drugs, pharmaceuticals,
 CC therapies and interventions which may be effective in treating
 CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
 CC disease. The animal models for trinucleotide repeat disorders are ideal
 CC model systems to study the progression of disease in vivo, the molecular
 CC basis of these diseases and show the features observed in human disease.
 CC using the mice, it is possible to model both the pathogenic mechanism and
 CC the trinucleotide repeat instability in the mouse.

XX Sequence 1839 BP; 424 A; 522 C; 535 G; 358 T; 0 other;
 SQ

Query Match 100.0%; Score 25; DB 22; Length 1839;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 |||||
 Db 74 AGCTCAGACATGGACTCCATGGCCC 50

RESULT 6
 AAS05286/c
 ID AAS05286 standard; DNA; 1848 BP.

XX AAS05286;
 AC
 DT 07-SEP-2001 (first entry)
 XX

DE DNA sequence of an expanded T243 gene.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; mouse; mutant; ds.

XX Mus sp.
 OS Synthetic.

XX Key Location/Qualifiers

FF CDS 55..894
 FT /*tag= a
 FT /product= "Expanded TRP"

XX WO200130798-A1.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-US29382.

XX 26-OCT-1999; 99US-0161488.

XX (DELT-) DELTAGEN INC.

XX Klein R, Matthews W, Moore M, Allen KD;

XX WPI; 2001-300473/31.

XX P-PSDB; AAU02500.

XX Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 PT contains heterozygous disruption in a gene encoding TRP -

XX Disclosure; Fig 15; 106pp; English.

XX The present sequence represents the DNA sequence of an expanded T243

CC gene. The T243 gene encodes, for a trinucleotide repeat protein (TRP).
 CC The invention describes methods of producing embryonic stem (ES)
 CC cells comprising a heterozygous disruption in a target DNA sequence
 CC (preferably gene T243) encoding a TRP and of producing a knockout mouse
 CC comprising a homozygous disruption in a gene encoding TRP, where the
 CC disruption inhibits the production of the wild type TRP. The invention
 CC also relates to identifying agents capable of affecting a phenotype of
 CC a knockout mouse. Also described are methods of determining whether
 CC expansion of the trinucleotide repeat in a gene encoding TRP produces a
 CC phenotypic change. The transgenic animals and the cells are useful for
 CC identifying compounds capable of ameliorating disease symptoms, and as
 CC test substrates for the identification of drugs, pharmaceuticals,
 CC therapies and interventions which may be effective in treating
 CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
 CC disease. The animal models for trinucleotide repeat disorders are ideal
 CC model systems to study the progression of disease in vivo, the molecular
 CC basis of these diseases and show the features observed in human disease.
 CC using the mice, it is possible to model both the pathogenic mechanism and
 CC the trinucleotide repeat instability in the mouse.

XX Sequence 1848 BP; 415 A; 519 C; 543 G; 371 T; 0 other;
 SQ

Query Match 100.0%; Score 25; DB 22; Length 1848;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 |||||
 Db 74 AGCTCAGACATGGACTCCATGGCCC 50

RESULT 7
 AAS21311/c
 ID AAS21311 standard; cDNA; 1675 BP.

XX AAS21311;
 AC
 DT 24-OCT-2001 (first entry)
 XX

DE Human cDNA sequence encoding for PRO4409 polypeptide.
 KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.

XX Homo sapiens.

XX WO200140466-A2.

XX 07-JUN-2001.

XX 01-DEC-2000; 2000WO-US32678.

XX 01-DEC-1999; 99WO-US28301.

XX 01-DEC-1999; 99WO-US28634.

XX 02-DEC-1999; 99WO-US28551.

XX 02-DEC-1999; 99WO-US28564.

XX 02-DEC-1999; 99WO-US28565.

XX 09-DEC-1999; 99US-0170262.

XX 16-DEC-1999; 99WO-US30095.

XX 20-DEC-1999; 99WO-US30999.

XX 30-DEC-1999; 99WO-US31243.

XX 06-JAN-2000; 2000WO-US00277.

XX 06-JAN-2000; 2000WO-US00376.

XX 11-FEB-2000; 2000WO-US03565.

XX 18-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04342.

XX 22-FEB-2000; 2000WO-US04414.

XX 24-FEB-2000; 2000WO-US04914.

XX 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.
 PA (GETH) GENENTECH INC.
 XX
 XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI: 2001-408281/43.
 DR P-PSDB; AAU12239.
 XX
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX
 PS Claim 3: Fig 135; 813pp; English.
 XX
 CC AAS21244-AAS21518 encode for novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 other;

Query Match 76.8%; Score 19.2; DB 22; Length 1675;
 Best Local Similarity 87.5%; Pred. No. 39;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gctcagacatggactccatggccc 25
 ||||| ||| ||||| |||||
 DB 355 GCTCAGCATGTAATCCATGCCCC 332

RESULT 8
 ID AAA27030/c
 XX AAA27030 standard; cDNA; 435 BP.
 AC AAA27030;
 XX
 DT 22-AUG-2000 (first entry)
 XX Human EST with GenBank Accession No. AAL12513.
 DE
 XX Human; cbl-SL; proto-oncogene; tyrosine kinase activity modulation;
 KW cancer; carcinoma; cell proliferation; cell differentiation;
 KW cytostatic; gene therapy; EST; expressed sequence tag; ss.
 XX

OS Homo sapiens.
 XX WO200027865-A1.
 PN
 XX 18-MAY-2000.
 PD
 XX
 PF 05-NOV-1999; 99WO-US26057.
 XX
 XX 06-NOV-1998; 98US-0107470.
 PR
 XX (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
 XX
 XX Borriello F, Band H;
 PI
 XX WPI: 2000-376482/32.
 DR
 XX Novel nucleic acids encoding cbl-SL proteins useful for treating
 PT cancers expressing a mutant cbl-SL polypeptide
 PT
 PS Claim 21; Page 71; 77pp; English.
 XX
 CC The present sequence is a human EST with Genbank Accession Number
 CC AAL12513. It shows partial homology to the coding sequence for the human
 CC protein cbl-SL (Y94330), which is believed to play a role in the
 CC regulation of a cell's growth, differentiation and proliferation.
 CC Aberrant expression levels and mutant forms of tyrosine kinases which are
 CC the likely targets of cbl-SL have been found in patients with certain
 CC types of cancer. The cbl-SL nucleotide sequence was isolated by screening
 CC a human pancreatic adenocarcinoma cDNA library using the insert of
 CC I.M.A.G.E clone ID 526956 from ATCC. Individual cDNA clones were
 CC subjected to 3-4 rounds of amplification and purification and several
 CC inserts were sequenced. The sequence of the resulting cbl-SL full-length
 CC clone was confirmed on more independent clones. The nucleic acid sequence
 CC may be used to screen for the presence of a carcinoma in a subject by
 CC comparing the cbl-SL DNA or mRNA sequence in the test sample to the
 CC cbl-SL DNA or mRNA of a control sample. Levels of cbl-1 expression may
 CC be measured using PCR and Northern blotting. The nucleic acid sequence
 CC may also be used in gene therapy and in the construction of a cbl-SL gene
 CC knockout in cells and animals used to study cbl-SL activity. cbl-SL
 CC polypeptides can be used to screen peptide libraries, including phage
 CC display libraries, to identify and select peptide binding partners of the
 CC cbl-SL protein. These can be used for screening assays, for purification
 CC protocols and for interfering directly with the functioning of cbl-SL.
 CC Any binding agents that also bind to the present sequence are excluded
 CC from the invention. The cbl-SL protein may also be administered as a
 CC therapeutic agent to cancer patients expressing a mutant cbl-SL.
 XX
 SQ Sequence 435 BP; 94 A; 136 C; 121 G; 82 T; 2 other;

Query Match 74.4%; Score 18.6; DB 21; Length 435;
 Best Local Similarity 84.0%; Pred. No. 63;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 ||||| ||| ||||| |||||
 DB 287 AGCTCAAAATGTGGAGTCCATGCCCC 263

RESULT 9
 ID AAS26338
 XX AAS26338 standard; cDNA; 585 BP.
 AC AAS26338;
 XX
 DT 07-NOV-2001 (first entry)
 XX Human cDNA encoding a novel secreted protein, Seq ID 517.
 DE
 XX Human; immunosuppressive; antiarthritic; ss; antirheumatic;
 KW cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
 KW vulnery; secreted protein; rheumatoid arthritis;

KW	hyperproliferative disorder; cardiovascular disorder; cardiac arrest;	14-SEP-2000;	2000US-0233065.
KW	cerebrovascular disorder; cerebral ischaemia; angiogenesis;	PR 21-SEP-2000;	2000US-0234223.
KW	nervous system disorder; Alzheimer's disease; infection; ocular disorder;	PR 21-SEP-2000;	2000US-0234274.
KW	corneal infection; wound healing; epithelial cell proliferation;	PR 25-SEP-2000;	2000US-0234997.
KW	skin ageing; food additive; preservative; antiproliferative.	PR 25-SEP-2000;	2000US-0234998.
XX		PR 26-SEP-2000;	2000US-0235484.
OS	Homo sapiens.	PR 27-SEP-2000;	2000US-0235834.
XX		PR 27-SEP-2000;	2000US-0235836.
XX	WO200155322-A2.	PR 29-SEP-2000;	2000US-0236327.
XX		PR 29-SEP-2000;	2000US-0236367.
XX		PR 29-SEP-2000;	2000US-0236368.
PD		PR 29-SEP-2000;	2000US-0236369.
XX		PR 29-SEP-2000;	2000US-0236370.
XX		PR 02-OCT-2000;	2000US-0236802.
XX		PR 02-OCT-2000;	2000US-0237037.
XX		PR 02-OCT-2000;	2000US-0237038.
PR		PR 02-OCT-2000;	2000US-0237039.
PR		PR 02-OCT-2000;	2000US-0237040.
PR		PR 13-OCT-2000;	2000US-0239935.
PR		PR 13-OCT-2000;	2000US-0239937.
PR		PR 20-OCT-2000;	2000US-0240960.
PR		PR 20-OCT-2000;	2000US-0241221.
PR		PR 20-OCT-2000;	2000US-0241785.
PR		PR 20-OCT-2000;	2000US-0241786.
PR		PR 20-OCT-2000;	2000US-0241787.
PR		PR 20-OCT-2000;	2000US-0241808.
PR		PR 20-OCT-2000;	2000US-0241809.
PR		PR 20-OCT-2000;	2000US-0241826.
PR		PR 01-NOV-2000;	2000US-0244617.
PR		PR 08-NOV-2000;	2000US-0246474.
PR		PR 08-NOV-2000;	2000US-0246475.
PR		PR 08-NOV-2000;	2000US-0246476.
PR		PR 08-NOV-2000;	2000US-0246477.
PR		PR 08-NOV-2000;	2000US-0246478.
PR		PR 08-NOV-2000;	2000US-0246523.
PR		PR 08-NOV-2000;	2000US-0246524.
PR		PR 08-NOV-2000;	2000US-0246525.
PR		PR 08-NOV-2000;	2000US-0246526.
PR		PR 08-NOV-2000;	2000US-0246527.
PR		PR 08-NOV-2000;	2000US-0246528.
PR		PR 08-NOV-2000;	2000US-0246532.
PR		PR 08-NOV-2000;	2000US-0246610.
PR		PR 08-NOV-2000;	2000US-0246611.
PR		PR 08-NOV-2000;	2000US-0246613.
PR		PR 17-NOV-2000;	2000US-0249207.
PR		PR 17-NOV-2000;	2000US-0249208.
PR		PR 17-NOV-2000;	2000US-0249209.
PR		PR 17-NOV-2000;	2000US-0249210.
PR		PR 17-NOV-2000;	2000US-0249211.
PR		PR 17-NOV-2000;	2000US-0249212.
PR		PR 17-NOV-2000;	2000US-0249213.
PR		PR 17-NOV-2000;	2000US-0249214.
PR		PR 17-NOV-2000;	2000US-0249215.
PR		PR 17-NOV-2000;	2000US-0249216.
PR		PR 17-NOV-2000;	2000US-0249217.
PR		PR 17-NOV-2000;	2000US-0249218.
PR		PR 17-NOV-2000;	2000US-0249244.
PR		PR 17-NOV-2000;	2000US-0249245.
PR		PR 17-NOV-2000;	2000US-0249264.
PR		PR 17-NOV-2000;	2000US-0249265.
PR		PR 17-NOV-2000;	2000US-0249297.
PR		PR 17-NOV-2000;	2000US-0249299.
PR		PR 17-NOV-2000;	2000US-0249300.
PR		PR 01-DEC-2000;	2000US-0250160.
PR		PR 01-DEC-2000;	2000US-0250391.
PR		PR 05-DEC-2000;	2000US-0251030.
PR		PR 05-DEC-2000;	2000US-0251988.
PR		PR 05-DEC-2000;	2000US-0256719.
PR		PR 06-DEC-2000;	2000US-0251479.
PR		PR 08-DEC-2000;	2000US-0251856.
PR		PR 08-DEC-2000;	2000US-0251868.
PR		PR 08-DEC-2000;	2000US-0251869.


```
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488783/53.
DR P-PSDB; AAU16351.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 1; SEQ ID No 517; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunoassorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.
Query Match 74.48; Score 18.6; DB 22; Length 585;
Best Local Similarity 84.08; Pred. No. 65;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 1 agctcagacatggactccatggccc 25
Db 484 agctcaaatgtgagtcacatggccc 508
RESULT 10
AAS26339/c
ID AAS26339 standard; cDNA; 933 BP.
XX
XX AAS26339;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human cDNA encoding a novel secreted protein, Seq ID 518.
XX
XX Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
XX skin aging; food additive; preservative; antiproliferative.
XX
XX Homo sapiens.
```

```
XX WO200155322-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01341.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220984.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
```

PR	27-SEP-2000;	2000US-0235836.	PI	Rosen CA, Barash SC, Ruben SM;
PR	29-SEP-2000;	2000US-0236327.	XX	
PR	29-SEP-2000;	2000US-0236367.	DR	WPI; 2001-488783/53:
PR	29-SEP-2000;	2000US-0236368.	DR	P-PSDB; AAU16352.
PR	29-SEP-2000;	2000US-0236369.	XX	
PR	29-SEP-2000;	2000US-0236370.	PT	New nucleic acid molecules encoding 461 human secreted proteins for
PR	29-SEP-2000;	2000US-0236370.	PT	diagnosing, preventing, treating or ameliorating medical conditions and
PR	02-OCT-2000;	2000US-0236802.	PT	used as food additives or preservatives -
PR	02-OCT-2000;	2000US-0237037.	XX	
PR	02-OCT-2000;	2000US-0237038.	PS	Claim 1; SEQ ID No 518; 980pp; English.
PR	02-OCT-2000;	2000US-0237039.	XX	
PR	02-OCT-2000;	2000US-0237040.	XX	The invention relates to isolated nucleic acid molecules and their
PR	13-OCT-2000;	2000US-0239935.	CC	encoded secreted proteins. The nucleic acids and proteins are used to
PR	13-OCT-2000;	2000US-0239935.	CC	prevent, treat or ameliorate a medical condition in e.g. humans, mice,
PR	13-OCT-2000;	2000US-0239937.	CC	rabbits, goats, horses, cats, dogs, chickens or sheep. They
PR	20-OCT-2000;	2000US-0240960.	CC	are also used in diagnosing a pathological condition or susceptibility
PR	20-OCT-2000;	2000US-0241221.	CC	to a pathological condition. Antibodies to the proteins can also
PR	20-OCT-2000;	2000US-0241785.	CC	be used in alleviating symptoms associated with the disorders and in
PR	20-OCT-2000;	2000US-0241786.	CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
PR	20-OCT-2000;	2000US-0241787.	CC	immunosorbant assays (ELISA). Disorders which are diagnosed or treated
PR	20-OCT-2000;	2000US-0241808.	CC	include autoimmune diseases e.g. rheumatoid arthritis,
PR	20-OCT-2000;	2000US-0241809.	CC	hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders
PR	20-OCT-2000;	2000US-0241826.	CC	e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
PR	01-NOV-2000;	2000US-0244617.	CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi
PR	08-NOV-2000;	2000US-0246474.	CC	and ocular disorders e.g. corneal infection, and many other
PR	08-NOV-2000;	2000US-0246475.	CC	disorders listed in the specification. The polypeptides can also
PR	08-NOV-2000;	2000US-0246476.	CC	be used to aid wound healing and epithelial cell proliferation, to
PR	08-NOV-2000;	2000US-0246477.	CC	prevent skin aging due to sunburn, to maintain organs before
PR	08-NOV-2000;	2000US-0246523.	CC	transplantation, for supporting cell culture of primary tissues, to
PR	08-NOV-2000;	2000US-0246524.	CC	regenerate tissues and in chemotaxis. The polypeptides can also be used
PR	08-NOV-2000;	2000US-0246525.	CC	as a food additive or preservative to increase or decrease storage
PR	08-NOV-2000;	2000US-0246526.	CC	capabilities, fat content, lipid, protein, carbohydrate, vitamins,
PR	08-NOV-2000;	2000US-0246527.	CC	minerals, cofactors and other nutritional components. The present
PR	08-NOV-2000;	2000US-0246528.	CC	sequence encodes a novel secreted protein of the invention.
PR	08-NOV-2000;	2000US-0246532.		
PR	08-NOV-2000;	2000US-0246609.		
PR	08-NOV-2000;	2000US-0246610.		
PR	08-NOV-2000;	2000US-0246611.		
PR	08-NOV-2000;	2000US-0246613.		
PR	17-NOV-2000;	2000US-0249207.		
PR	17-NOV-2000;	2000US-0249208.		
PR	17-NOV-2000;	2000US-0249209.		
PR	17-NOV-2000;	2000US-0249210.		
PR	17-NOV-2000;	2000US-0249211.		
PR	17-NOV-2000;	2000US-0249212.		
PR	17-NOV-2000;	2000US-0249213.		
PR	17-NOV-2000;	2000US-0249214.		
PR	17-NOV-2000;	2000US-0249215.		
PR	17-NOV-2000;	2000US-0249216.		
PR	17-NOV-2000;	2000US-0249217.		
PR	17-NOV-2000;	2000US-0249218.		
PR	17-NOV-2000;	2000US-0249219.		
PR	17-NOV-2000;	2000US-0249245.		
PR	17-NOV-2000;	2000US-0249264.		
PR	17-NOV-2000;	2000US-0249265.		
PR	17-NOV-2000;	2000US-0249297.		
PR	17-NOV-2000;	2000US-0249299.		
PR	17-NOV-2000;	2000US-0249300.		
PR	01-DEC-2000;	2000US-0250160.		
PR	01-DEC-2000;	2000US-0250391.		
PR	05-DEC-2000;	2000US-0251030.		
PR	05-DEC-2000;	2000US-0251388.		
PR	05-DEC-2000;	2000US-0256719.		
PR	06-DEC-2000;	2000US-0251479.		
PR	08-DEC-2000;	2000US-0251856.		
PR	08-DEC-2000;	2000US-0251856.		
PR	08-DEC-2000;	2000US-0251869.		
PR	08-DEC-2000;	2000US-0251889.		
PR	08-DEC-2000;	2000US-0251990.		
PR	11-DEC-2000;	2000US-0254097.		
PR	05-JAN-2001;	2001US-0259678.		
PA	(HUMA-)	HUMAN GENOME SCI INC.		
XX				

Query Match

74.4%; Score 18.6; DB 22; Length 933;

Best Local Similarity

84.0%; Pred: No. 68;

Matches

21; Conservative

0; Mismatches

4; Indels

0; Gaps

0;

QY:

1 agctcagacatggactccatggccc 25

||||| | ||| |||||

DB

447 AGCTCAATGTGGAGTCCATGSCC 423

RESULT 11

AAA27028/c

ID

AAA27028 standard; cDNA: 1422 BP.

XX

AAA27028;

AC

AAA27028;

XX

22-AUG-2000 (first entry)

DE

Human cbl-SL cDNA largest open reading frame.

XX

Human; proto-oncogene; tyrosine kinase activity modulation;

KW

cancer; carcinoma; cell proliferation; cell differentiation;

KW

cytostatic; gene therapy; ss.

XX

Homo sapiens.

OS

Homo sapiens.

XX

FH

Key

Location/Qualifiers

FT

CDS

1..1422

FT

/*tag= a

FT

/product= "cbl-SL"

XX

XX

WO2000027865-A1.

XX

XX

18-MAY-2000.

PD

XX

05-NOV-1999;

99WO-US26057.

XX

XX

06-NOV-1998;

98US-0107470.

PR

XX PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
 XX PI Borriello F, Band H;
 XX PF WPI: 2000-376482/32.
 DR DR P-PSDB: AAY94330.
 XX
 PT Novel nucleic acids encoding cbl-SL proteins useful for treating
 XX cancers expressing a mutant cbl-SL polypeptide
 XX
 PS Claim 3; Page 70-71; 77pp; English.
 XX
 CC The present sequence is the largest open reading frame of the human
 CC cbl-SL cDNA (AAA27027) which encodes the human protein cbl-SL. cbl-SL
 CC is believed to play a role in the regulation of a cell's growth,
 CC differentiation and proliferation. Aberrant expression levels and mutant
 CC forms of tyrosine kinases which are the likely targets of cbl-SL have
 CC been found in patients with certain types of cancer. The sequence was
 CC isolated by screening a human pancreatic adenocarcinoma cDNA library
 CC using the insert of I.M.A.G. E clone ID 526956 from ATCC. Individual cDNA
 CC clones were subjected to 3-4 rounds of amplification and purification
 CC and several inserts were sequenced. The sequence of the resulting cbl-SL
 CC full-length clone was confirmed on more independent clones. The
 CC nucleic acid sequence may be used to screen for the presence of a
 CC carcinoma in a subject by comparing the cbl-SL DNA or mRNA sequence in
 CC the test sample to the cbl-SL DNA or mRNA of a control sample. Levels of
 CC cbl-1 expression may be measured using PCR and Northern blotting. The
 CC nucleic acid sequence may also be used in gene therapy and in the
 CC construction of a cbl-SL gene knockout in cells and animals used to
 CC study cbl-SL activity. cbl-SL polypeptides can be used to screen peptide
 CC libraries, including phage display libraries, to identify and select
 CC peptide binding partners of the cbl-SL protein. These can be used
 CC for screening assays, for purification protocols and for interfering
 CC directly with the functioning of cbl-SL. The protein may also
 CC be administered as a therapeutic agent to cancer patients expressing a
 CC mutant cbl-SL.
 XX
 SQ Sequence 1422 BP; 269 A; 484 C; 432 G; 237 T; 0 other;

Query Match 74.4%; Score 18.6; DB 21; Length 1422;
 Best Local Similarity 84.0%; Pred. No. 71;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatgccc 25
 ||||| ||||| ||||| |||||
 Db 1049 AGCTCAAAATGTGGAGTCATGCGCC 1025

RESULT 12
 AAS25883/C
 ID AAS25883 standard; cDNA; 1522 BP.
 XX AC
 XX AAS25883;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human cDNA encoding a novel secreted protein, Seq ID 62.
 KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
 KW cytosolic; cardiac; vasotropic; cerebroprotective; nontropic;
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
 KW vulnary; secreted protein; rheumatoid arthritis;
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
 KW corneal infection; wound healing; epithelial cell proliferation;
 KW skin ageing; food additive; preservative; antiproliferative.
 XX Homo sapiens.
 OS
 XX WO200155322-A2.
 PN

XX PD 02-AUG-2001.
 XX
 XX PF 17-JAN-2001; 2001WO-US01341.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225286.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 25-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.


```
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR P-PSDB: AAM78486.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 889-891; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1526 BP; 303 A; 505 C; 465 G; 253 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 1526;
Best Local Similarity 84.0%; Pred. No. 72;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 agctcagacatggagctccatggccc 25
Db 1049 AGCTCAAAATGTGGAGTCCATGCCCC 1025

RESULT 14
AAK52603/C
ID AAK52603 standard; cDNA; 1543 BP.
XX
XX AAK52603;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human polynucleotide SEQ ID NO 2132.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157190-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001WO-US04098.
PF
XX
XX 03-FEB-2000; 2000US-0496914.
PR
XX
XX 27-APR-2000; 2000US-0560875.
PR
XX
XX 20-JUN-2000; 2000US-0598075.
PR
XX
XX 19-JUL-2000; 2000US-0620325.
PR
XX
XX 01-SEP-2000; 2000US-0654936.
PR
XX
XX 15-SEP-2000; 2000US-0663561.
PR
XX
XX 20-OCT-2000; 2000US-0693325.
PR

PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR P-PSDB: AAM79470.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 4504; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1543 BP; 306 A; 515 C; 470 G; 252 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 1543;
Best Local Similarity 84.0%; Pred. No. 72;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 agctcagacatggagctccatggccc 25
Db 1066 AGCTCAAAATGTGGAGTCCATGCCCC 1042

RESULT 15
AAA27027/C
ID AAA27027 standard; cDNA; 1547 BP.
XX
XX AAA27027;
AC
XX
XX 22-AUG-2000 (first entry)
DT
XX
XX Human cbl-SL cDNA.
DE
XX
XX Human; proto-oncogene; tyrosine kinase activity modulation;
KW cancer; carcinoma; cell proliferation; cell differentiation;
KW cytostatic; gene therapy; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 13..1437
CDS
XX
XX /*tag= a
XX /product= "cbl-SL"
XX
XX WO200027865-A1.
PN
XX
XX 18-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26057.
PF
XX
XX 06-NOV-1998; 98US-0107470.
PR
XX
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
PA
```

Borriello F, Band H;
 WPI; 2000-376482/32.
 P-PSDB; AAY94330.
 Novel nucleic acids encoding cbl-SL proteins useful for treating
 cancers expressing a mutant cbl-SL polypeptide -
 Claim 2; Page 67-69; 77pp; English.
 The present sequence encodes the human protein cbl-SL, which is
 believed to play a role in the regulation of a cell's growth,
 differentiation and proliferation. Aberrant expression levels and mutant
 forms of tyrosine kinases which are the likely targets of cbl-SL have
 been found in patients with certain types of cancer. The sequence was
 isolated by screening a human pancreatic adenocarcinoma cDNA library
 using the insert of I.M.A.G.E clone ID 526956 from ATCC. Individual cDNA
 clones were subjected to 3-4 rounds of amplification and purification and
 several inserts were sequenced. The sequence of the resulting cbl-SL
 full-length clone was confirmed on more independent clones. The
 nucleic acid sequence may be used to screen for the presence of a
 carcinoma in a subject by comparing the cbl-SL DNA or mRNA sequence in
 the test sample to the cbl-SL DNA or mRNA of a control sample. Levels of
 cbl-1 expression may be measured using PCR and Northern blotting. The
 nucleic acid sequence may also be used in gene therapy and in the
 construction of a cbl-SL gene knockout in cells and animals used to
 study cbl-SL activity. cbl-SL polypeptides can be used to screen peptide
 libraries, including phage display libraries, to identify and select
 peptide binding partners of the cbl-SL protein. These can be used
 for screening assays, for purification protocols and for interfering
 directly with the functioning of cbl-SL. The protein may also
 be administered as a therapeutic agent to cancer patients expressing a
 mutant cbl-SL.

```
Query Match          74.4%; Score 18.6; DB 21; Length 1547;
Best Local Similarity 84.0%; Pred. NO. 72;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Qy 1 agctcagacatggactcccatgccc 25
||||| | ||||| |||||
1061 AGCTCAAAATGTGGAGTCCATGGCCC 1037

Search completed: May 24, 2002, 00:23:09
Job time: 9006 sec

CLASSIFICATION: 435
PRIOR APPLICATION DATA: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-913-805A-9

Query Match 65.6%; Score 16.4; DB 3; Length 2100;
Best Local Similarity 94.4%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 gctcagacatggactcca 19
Db 1294 GCTCTGACATGGACTCCA 1277

RESULT 15
US-09-442-629-9/c
Sequence 9, Application US/09442629
Patent No. 6291219
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
SHIBATA, Tetsuo
YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442.629
FILING DATE: 18-NO. 6291219-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/913.805A
FILING DATE: 7 JAN 1998
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996

APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-442-629-9

Query Match 65.6%; Score 16.4; DB 4; Length 2100;
Best Local Similarity 94.4%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 gctcagacatggactcca 19
Db 1294 GCTCTGACATGGACTCCA 1277

Search completed: May 24, 2002, 00:18:05
Job time: 11551 sec

```

; NAME/KEY: CDS
; LOCATION: (757)..(3771)
US-08-935-450-1

Query Match 67.2%; Score 16.8; DB 2; Length 4534;
Best Local Similarity 90.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 cagacatggactccatgcc 24
||||| ||||| ||||| |||||
Db 2686 CAGATGCTGCTCCATGCC 2667

RESULT 13
US-09-239-909-3
; Sequence 3, Application US/09239909
; Patent No. 6284952
; GENERAL INFORMATION:
; APPLICANT: Kumbo Petrochemical Co. Ltd.
; TITLE OF INVENTION: Transgenic Plants with Divergent ScaM4 or ScaM5 Gene to
; FILE REFERENCE: P99-2-6
; CURRENT APPLICATION NUMBER: US/09/239,909
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: EP 99300136.1
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: KOPATIN 1.0
; SEQ ID NO 3
; LENGTH: 916
; TYPE: DNA
; ORGANISM: G. max calmodulin5 (ScaM5)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (69)..(518)
US-09-239-909-3

Query Match 66.4%; Score 16.6; DB 4; Length 916;
Best Local Similarity 82.6%; Pred. No. 62;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatgcc 23
||||| ||||| || |||||
Db 51 agctcaaacattaacaccatgcc 73

RESULT 14
US-08-913-805A-9/C
; Sequence 9, Application US/08913805A
; Patent No. 6054304
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: UOZUMI, Naofumi
; APPLICANT: SHIBA, Tetsuo
; APPLICANT: YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998

```



```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 15079 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORGANISM: Streptomyces clavuligerus
; US-09-385-028-1

```

```

Query Match          68.08; Score 17; DB 4; Length 15079;
Best Local Similarity 80.0%; Pred. No. 60;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 agctcagacatggactgcctccat 25
    ||||| ||||| ||||| |||||
Db 3623 AGCTCCGACCTGGAGACCATCGCCC 3647

```

```

RESULT 9
US-08-322-742-18
; Sequence 18, Application US/08322742
; Patent No. 5688641
; GENERAL INFORMATION:
; APPLICANT: Sager, Ruth
; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,742
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/938,823
; FILING DATE: September 1, 1992
; APPLICATION NUMBER: 07/844,296
; FILING DATE: February 28, 1992
; APPLICATION NUMBER: 07/552,216
; FILING DATE: February 28, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/048003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 539
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-322-742-18

```

```

Query Match          67.2%; Score 16.8; DB 1; Length 539;
Best Local Similarity 90.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 agctcagacatggactgcctccat 20

```

```

Db 356 AACTCAGACATGGACTACAT 375
    | ||||| ||||| ||||| |||||
RESULT 10
US-08-923-454A-17
; Sequence 17, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Livi, George
; APPLICANT: Karran, Eric
; APPLICANT: Clinkenbeard, Helen
; APPLICANT: Browne, Michael
; APPLICANT: Southan, Christopher
; TITLE OF INVENTION: HUMAN SERINE PROTEASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,454A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025436
; FILING DATE: 06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2036 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORIGINAL SOURCE: Feature polymorphism at 1325
US-08-923-454A-17

```

```

Query Match          67.2%; Score 16.8; DB 3; Length 2036;
Best Local Similarity 90.0%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 agctcagacatggactgcctccat 20
    | ||||| ||||| ||||| |||||
Db 979 AACTCAGACATGGACTACAT 998

```

```

RESULT 11
US-08-888-077A-41
; Sequence 41, Application US/08888077A
; Patent No. 6020143
; GENERAL INFORMATION:

```

```

? APPLICATION NUMBER: US/09/385,028
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? PRIORITY NUMBER: US 08/790,462
? FILING DATE: 29-JAN-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: D. Douglas Price
? REGISTRATION NUMBER: 24,514
? REFERENCE/DOCKET NUMBER: 1418/P57452US2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 638-6666
? TELEFAX: (202) 39305350
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 11604 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-09-385-028-13

Query Match 68.0%; Score 17; DB 4; Length 11604;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatgccc 25
||||| ||||| ||||| ||||| |||||
DB 1591 AGCTCGACCTGGAGCATTGCCCC 1615

RESULT 8
US-09-385-028-1
? Sequence 1, Application US/09385028
? Patent No. 6232106
? GENERAL INFORMATION:
? APPLICANT: Susan E. Jensen
? APPLICANT: Kwamena A Aidoo
? APPLICANT: Ashish S. Paradkar
? TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
? Patent No. 6232106
? TITLE OF INVENTION: Acid Biosynthesis
? NUMBER OF SEQUENCES: 25
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN,PPLC
? STREET: The Jennifer Buliding, 400 Seventh Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.A.
? ZIP: 20004
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/385,028
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? PRIOR APPLICATION NUMBER: US 08/790,462
? FILING DATE: 29-JAN-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: D. Douglas Price
? REGISTRATION NUMBER: 24,514
? REFERENCE/DOCKET NUMBER: 1418/P57452US2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 638-6666
? TELEFAX: (202) 39305350
? TELEX: RCA 248593 IDEA UR
? INFORMATION FOR SEQ ID NO: 1:

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 21:05:34 ; Search time 121.39 Seconds
(without alignments)
50.588 Million cell updates/sec

Title: US-09-696-686-45
Perfect score: 25
Sequence: 1 agctcagacatggactccatggccc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	17.6	70.4	3228	4	US-09-318-448-20
2	17.6	70.4	7720	4	US-09-318-448-5
3	17.2	68.8	4403765	4	US-09-103-840A-2
4	17.2	68.8	4411529	4	US-09-103-840A-1
5	17	68.0	1722	4	US-09-385-028-15
6	17	68.0	2925	2	US-09-014-969-10
7	17	68.0	11604	4	US-09-385-028-13
8	17	68.0	15079	4	US-09-385-028-1
9	16.8	67.2	539	1	US-08-322-742-18
10	16.8	67.2	2036	3	US-08-923-454A-17
11	16.8	67.2	2205	3	US-08-888-077A-41
12	16.8	67.2	4534	2	US-08-935-450-1
13	16.6	66.4	916	4	US-09-239-909-3
14	16.4	65.6	2100	3	US-08-913-805A-9
15	16.4	65.6	2100	4	US-09-442-629-9
16	16.2	64.8	2268	2	US-08-890-853-1
17	16.2	64.8	2268	2	US-09-099-125A-1
18	16.2	64.8	2268	2	US-09-099-124A-1
19	16.2	64.8	2268	2	US-09-197-008-1
20	16.2	64.8	2268	4	US-09-032-476-1
21	16.2	64.8	2268	4	US-08-890-854-1
22	16.2	64.8	2268	4	US-09-023-324-1
23	16.2	64.8	2271	4	US-08-910-820-8
24	16.2	64.8	2931	4	US-09-168-629-14
25	16.2	64.8	3966	3	US-09-215-131-1
26	16.2	64.8	3966	3	US-09-222-734-1
27	16.2	64.8	7886	2	US-08-751-189-2

C 28	16.2	64.8	7886	2	US-09-060-836-2	Sequence 2, Appli
C 29	16.2	64.8	7886	4	US-09-184-445-2	Sequence 2, Appli
C 30	16	64.0	327	4	US-08-836-075A-5	Sequence 5, Appli
C 31	16	64.0	808	3	US-08-651-136C-15	Sequence 15, Appli
C 32	16	64.0	2415	3	US-09-019-689-1	Sequence 1, Appli
C 33	16	64.0	3407	1	US-08-253-155A-7	Sequence 7, Appli
C 34	16	64.0	4374	2	US-08-449-644-3	Sequence 3, Appli
C 35	16	64.0	4374	2	US-08-087-244A-3	Sequence 3, Appli
C 36	16	64.0	11531	1	US-08-068-945A-1	Sequence 1, Appli
C 37	16	64.0	11531	1	US-08-442-806-1	Sequence 1, Appli
C 38	15.8	63.2	1329	2	US-08-758-621-1	Sequence 1, Appli
C 39	15.8	63.2	1329	4	US-09-107-858-1	Sequence 46, Appli
C 40	15.8	63.2	2464	4	US-09-268-992-46	Sequence 48, Appli
C 41	15.8	63.2	2488	4	US-09-268-992-48	Sequence 47, Appli
C 42	15.8	63.2	3016	4	US-09-268-992-47	Sequence 246, Appli
C 43	15.8	63.2	3323	4	US-08-936-165A-246	Sequence 20, Appli
C 44	15.6	62.4	815	4	US-09-541-941B-20	Sequence 7, Appli
C 45	15.6	62.4	877	3	US-08-545-809A-7	

ALIGNMENTS

RESULT 1
US-09-318-448-20
; Sequence 20, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318.448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 3228
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-20

Query Match 70.4%; Score 17.6; DB 4; Length 3228;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gctcagacatggactccatggccc 25
||||| || || ||||| ||
Db 2785 gctcagcaggccctccatggccc 2808

RESULT 2
US-09-318-448-5
; Sequence 5, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318.448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-5

```

FEATURES      High quality sequence stop: 262.
source        Location/Qualifiers
1..651
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone_lib="IMAGE:3484616"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT   163 a   212 c   172 g   104 t
ORIGIN

Query Match      100.0%; Score 25; DB 10; Length 651;
Best Local Similarity 100.0%; Pred. NO. 4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 agctcagacatggactccatggccc 25
    |
Db  98 AGCTCAGACATGGACTCCATGGCCC 74

```

Search completed: May 23, 2002, 23:05:27
Job time: 7483 sec

/note="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGGCGCGCACTCGAGTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAAATTAATTCCTCCCGCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

138 a 162 c 216 g 127 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 643;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agctcagacatggactccatggccc 25
|||||

Db 101 AGCTCAGACATGGACTCCATGGCCC 77

RESULT 14
BB652011/c
LOCUS
DEFINITION BB652011 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330027E02 5', mRNA sequence.
ACCESSION BB652011 GI:15402706
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Tel: 81-45-503-9216
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Funct. Genomics 2 pre, L72-L86 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES
source
1. 649
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C330027E02"
/clone_lib="RIKEN full-length enriched, ES cells"
/cell_type="ES cells"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGGATCCAGAGCTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAAATTAATTCCTCCCGCCCCC 3']"

BASE COUNT 135 a 170 c 197 g 147 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 649;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agctcagacatggactccatggccc 25
|||||

Db 94 AGCTCAGACATGGACTCCATGGCCC 70

RESULT 15
BE289316/c
LOCUS
DEFINITION BE289316 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484616 5', mRNA sequence.
ACCESSION BE289316 GI:9169606
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsbbs-r@mail.nih.gov
cDNA Library Preparation: Lohar Hennighausen Ph.D., Robin Humphreys
Tissue Procurement: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM8517 row: n column: 09

D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

TITLE JOURNAL COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES source

Location/Qualifiers
1. .641
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A530052E13"
/clone_lib="RIKEN full-length enriched, adult male aorta
and vein"
/sex="male"
/tissue_type="aorta and vein"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGTCACAGAGCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5' GAGAGAGAGATTCGAGTAATTAATTAATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 142 a 161 c 206 g 132 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 641;
Best Local Similarity 100.0%; Pred. No. 4;

Matches

25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agctcagacatggactccatgcc 25
|||||
Db 67 AGCTCAGACATGGACTCCATGCC 43

RESULT 13

BB661603/C
LOCUS BB661603 643 bp mRNA linear EST 26-OCT-2001
DEFINITION BB661603 RIKEN full-length enriched, 0 day neonate kidney Mus
musculus cDNA clone D630047M13 5', mRNA sequence.

ACCESSION BB661603
VERSION BB661603
KEYWORDS EST
SOURCE house mouse.

ORGANISM Mus musculus
REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse tissues.
Location/Qualifiers
1. .643
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630047M13"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"

FEATURES
source
Location/Qualifiers
1. .643
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630047M13"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"

Human Genome Sequences. Mamm. Genome 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse tissues.
Location/Qualifiers
1. .643
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630047M13"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"

```

/tissue_type="spinal cord"
/dev_stage="adult"
/lab_host="PH10B"
/notes="Site.1: Sali; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
GAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATCTCGATTAAATTAATCCCTCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
BASE COUNT      145 a 152 c 200 g 128 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 625;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
|||||
Db 37 AGCTCAGACATGGACTCCATGGCCC 13

RESULT 11
BB650071/c
LOCUS
DEFINITION BB650071 RIKEN full-length enriched, 0 day neonate cerebellum Mus
musculus cDNA clone C230039G09 5', mRNA sequence.
ACCESSION BB650071
VERSION BB650071.1 GI:16484326
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 637)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Sano,H.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
source
1. 637
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, 0 day neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="PH10B"
/notes="Site.1: Sali; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
GAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 479.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATCTCGATTAAATTAATCCCTCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
BASE COUNT      140 a 164 c 220 g 109 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 637;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
|||||
Db 93 AGCTCAGACATGGACTCCATGGCCC 69

RESULT 12
BB637089/c
LOCUS
DEFINITION BB637089 RIKEN full-length enriched, adult male aorta and vein Mus
musculus cDNA clone A530052E13 5', mRNA sequence.
ACCESSION BB637089
VERSION BB637089.1 GI:16472874
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 641)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

```


Query Match	100.0%;	Score 25;	DB 9;	Length 623;
Best Local Similarity	100.0%;	Pred. NO. 4;		

Qy	1	agctcagacatggactccatgccc	25
Dd	89	AGCTCAGACATGGACTCCATGCCCC	65

RESULT_10	BB6333565/c	625 bp	linear	EST 26-OCT-2001
LOCUS	BB6333565/c			
DEFINITION	musculus cDNA clone A33062B19 5', mRNA sequence.			
ACCESSION	BB6333565			
VERSION	BB6333565.1			

ORGANISM	Mus musculus
house mouse	

REFERENCE
 AUTHORS
 ARAKAWA, T. 1965. *Ichuensis*: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 1. *Basia*. CO 625.
 ARKAT, C., and J. P. FUKUDA, S., FURUNO, M., HANAGAKI, T., HARA, A.,
 KAWAI, K., MORI, F., TSHI, Y., ITO, M., KAWAI, J., KONNO, H., KOUDA
 M., KOYA, S., MATSUYAMA, T., MIYAZAKI, A., NOMURA, K., OHNO, M.,
 OKAZAKI, Y., OKIDOMI, S., SAKAI, K., SAKAI, K., SANO, H., SASAKI
 D., SHIBATA, K., SHINGA, S., SUGI, T., SUGI, T., SUGI, T., SUZUKI, H.,
 TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TAKEDA, Y., TANAKA, T.,
 MURAMATSU, M. and HAYASHI, T.

TITLE
JOURNAL
COMMENT

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

```

/db_xref="taxon:10090"
/clone="A330062B19"
/clone_lib="RIKEN full-length enriched, adult male spinal
cord"
/sex="male"

```

Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 cDNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM9808 row: h column: 06
 High quality sequence start: 2
 High quality sequence stop: 599.

FEATURES

Location/Qualifiers

1..599

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4222133"

/clone_lib="NCI_CGAP_Kid14"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: kidney; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"

121 a.. 161 c 199 g 118 t

BASE COUNT

FEATURES source

Query Match

Best Local Similarity 100.0%; Score 25; DB 10; Length 599;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 agctcagacatgactccatgccc 25

|||||

Db 110 AGCTCAGACATGGACTCCATGGCCC 86

RESULT 8

BB605812/c

LOCUS

DEFINITION BB605812 RIKEN full-length enriched, 0 day neonate lung Mus

musculus cDNA clone E030013D02 5', mRNA sequence.

BB605812 608 bp mRNA linear EST 26-OCT-2001

BB605812.2 GI:16450910

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 608)

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.

, Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda

, Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,

Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki

,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,

Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Dec 5, 2000 this sequence version replaced gi:11557214.

Contact: Yoshinori Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: <http://genome-gsc.riken.go.jp/>

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 ,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
 encyclopedia: Real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
 ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
 Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome-gsc.riken.go.jp/>) for
 further details.

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Location/Qualifiers

1..608

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="E030013D02"

/clone_lib="RIKEN full-length enriched, 0 day neonate

lung"

/tissue_type="lung"

/dev_stage="0 day neonate"

/lab_host="DH10B"

/note="Site: 1: SalI; Site: 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGCGCGCAGCTCGAGTGTGTTTTTTTTTTTNN 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5'

GAGAGAGATCTCGAGTTAAATAATATCCGCCCCCCCC 3']. cDNA

was cleaved with BamHI and XhoI. Vector: a modified

pBluescript KS(+) after bulk excision from Lambda FLC I."

139 a 148 c 192 g 129 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 608;

Best Local Similarity 100.0%; Pred. No. 4;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatgactccatgccc 25

|||||

Db 38 AGCTCAGACATGGACTCCATGGCCC 14

TITLE

JOURNAL

COMMENT

RESULT 9

BB661026/c

LOCUS

DEFINITION

BB661026 RIKEN full-length enriched, 0 day neonate kidney Mus

musculus cDNA clone D630017H14 5', mRNA sequence.

623 bp mRNA linear EST 26-OCT-2001

BB661026

musculus

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 JOURNAL RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 COMMENT Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.

FEATURES Location/Qualifiers
 source
 1. .485
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370047K04"
 /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 BASE COUNT 101 a 132 c 154 g 98 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 9; Length 485;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 agctcagacatggactccatggccc 25
 ||||||||||||||||||||
 Db 92 AGCTCAGACATGGACTCCATGGCCC 68

RESULT 6
 BB853356/c
 LOCUS
 DEFINITION BB853356 RIKEN full-length enriched, B16 F10Y cells Mus musculus
 cDNA clone G370012D12 5', mRNA sequence.
 BB853356
 ACCESSION
 VERSION
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 502)
 REFERENCES
 AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
 Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
 Salto,R., Sakai,K., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
 Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa

A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
 Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)
 Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 source
 1. .502
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370012D12"
 /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 BASE COUNT 103 a 136 c 161 g 102 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 9; Length 502;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 agctcagacatggactccatggccc 25
 ||||||||||||||||||||
 Db 89 AGCTCAGACATGGACTCCATGGCCC 65
 RESULT 7
 BF780486/c
 LOCUS
 DEFINITION BF780486 599 bp mRNA linear EST 12-JAN-2001
 5', mRNA sequence.
 BF780486
 ACCESSION
 VERSION
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 599)
 REFERENCES
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-f@mail.nih.gov

JOURNAL COMMENT
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1..483
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"
CRL-2065 MLTC-1 cDNA
/cell_type="Leydig cells"
/cell_line="CRL-2065 MLTC-1"
BASE COUNT 92 a 127 c 156 g 97 t
ORIGIN

FEATURES Location/Qualifiers
source 1..483
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"
CRL-2065 MLTC-1 cDNA
/cell_type="Leydig cells"
/cell_line="CRL-2065 MLTC-1"
BASE COUNT 92 a 127 c 156 g 97 t
ORIGIN

FEATURES Location/Qualifiers
source 1..483
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"
CRL-2065 MLTC-1 cDNA
/cell_type="Leydig cells"
/cell_line="CRL-2065 MLTC-1"
BASE COUNT 92 a 127 c 156 g 97 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
|||||
DB 96 AGCTCAGACATGGACTCCATGGCCC 72

RESULT 4
BB857917/c
LOCUS BB857917 RIKEN full-length enriched, B16 F10Y cells EST 26-NOV-2001
DEFINITION CDNA clone G370047K11 5', mRNA sequence.
ACCESSION BB857917
VERSION BB857917.1 GI:17099371
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 483)

REFERENCE AUTHORS
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii
Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

JOURNAL COMMENT
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1..472
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, Leydig cells"
CRL-2065 MLTC-1 cDNA
/cell_type="Leydig cells"
/cell_line="CRL-2065 MLTC-1"
BASE COUNT 92 a 127 c 156 g 97 t
ORIGIN

FEATURES Location/Qualifiers
source 1..472
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, Leydig cells"
CRL-2065 MLTC-1 cDNA
/cell_type="Leydig cells"
/cell_line="CRL-2065 MLTC-1"
BASE COUNT 92 a 127 c 156 g 97 t
ORIGIN

FEATURES Location/Qualifiers
source 1..472
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, Leydig cells"
CRL-2065 MLTC-1 cDNA
/cell_type="Leydig cells"
/cell_line="CRL-2065 MLTC-1"
BASE COUNT 92 a 127 c 156 g 97 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
|||||
DB 96 AGCTCAGACATGGACTCCATGGCCC 72

RESULT 4
BB857917/c
LOCUS BB857917 RIKEN full-length enriched, B16 F10Y cells EST 26-NOV-2001
DEFINITION CDNA clone G370047K11 5', mRNA sequence.
ACCESSION BB857917
VERSION BB857917.1 GI:17099371
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 483)

REFERENCE AUTHORS
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii
Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES
SOURCE
 1. 260
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="E330035012"
 /clone_lib="RIKEN full-length enriched, 2 days pregnant adult female ovary"
 /sex="female"
 /tissue_type="ovary"
 /dev_stage="2 days pregnant adult"
 /lab_host="DH10B"
 /note="Site 1: Salt; Site 2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCGCAACTCGAGTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAATAATTAATCCCGCCCGCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I." 33 a 81 c 88 g 58 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 260;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 |||||
 Db 79 AGCTCAGACATGGACTCCATGGCCC 55

RESULT 2
AA002532/c
LOCUS
DEFINITION 406 bp mRNA linear EST 19-JUL-1996
 clone IMAGE:427563 5', mRNA sequence.
ACCESSION AA002532
VERSION AA002532.1 GI:1445971
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 406)
REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,N., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and Waterston,R.

TITLE **JOURNAL** **COMMENT**

The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:262115

Seq primer: mob.REGA+ET

High quality sequence stop: 328.

FEATURES **SOURCE**

1. 406
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:427563"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGCGCGCGCAATTTTTTTTTTTTTTTTTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru KO, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo." 81 a 107 c 128 g 90 t

BASE COUNT **ORIGIN**

Query Match 100.0%; Score 25; DB 9; Length 406;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 |||||
 Db 44 AGCTCAGACATGGACTCCATGGCCC 20

RESULT 3
BB863087/c
LOCUS
DEFINITION 472 bp mRNA linear EST 27-NOV-2001
 cDNA Mus musculus clone G430034D12 5', mRNA sequence.
ACCESSION BB863087
VERSION BB863087.1 GI:17109295
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 472)
REFERENCE
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinaawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akashira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 23, 2002, 21:00:44 ; Search time 4203.75 Seconds
(without alignments)
80.267 Million cell updates/sec

Title: US-09-696-686-45
Perfect score: 25
Sequence: 1 agctcagacatgactccatgccc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25	100.0	260	9	BB666889
C 2	25	100.0	406	9	AA002532
C 3	25	100.0	472	9	BB863087
C 4	25	100.0	483	9	BB857917
C 5	25	100.0	485	9	BB857913
C 6	25	100.0	502	9	BB853356
C 7	25	100.0	599	10	BF780486
C 8	25	100.0	608	9	BB605812
C 9	25	100.0	623	9	BB661026
C 10	25	100.0	625	9	BB633565
C 11	25	100.0	637	9	BB650071
C 12	25	100.0	641	9	BB637089
C 13	25	100.0	643	9	BB661603
C 14	25	100.0	649	9	BB652011
C 15	25	100.0	651	10	BE289316
C 16	25	100.0	776	10	BG864846
C 17	25	100.0	779	10	BI688404

C 18	25	100.0	793	10	BI104487
C 19	25	100.0	821	10	BI653380
C 20	25	100.0	838	10	BF534535
C 21	25	100.0	1553	11	AK005532
C 22	25	100.0	1586	11	AK010702
C 23	23.4	93.6	177	9	BB566317
C 24	23.4	93.6	223	9	BB578528
C 25	23.4	93.6	304	9	BB601182
C 26	23.4	93.6	382	9	AA048794
C 27	23.4	93.6	725	10	BG296771
C 28	22	88.0	495	10	BG062828
C 29	20.2	80.8	220	9	BB591610
C 30	20.2	80.8	330	9	BB856691
C 31	19.8	79.2	775	10	BG935945
C 32	19.2	76.8	253	9	AW889975
C 33	19.2	76.8	374	10	BF896723
C 34	19.2	76.8	408	10	BG407897
C 35	19.2	76.8	416	9	AW245217
C 36	19.2	76.8	419	9	AW406598
C 37	19.2	76.8	443	10	BG234205
C 38	19.2	76.8	448	9	AW630286
C 39	19.2	76.8	449	10	BE305036
C 40	19.2	76.8	471	9	AI878843
C 41	19.2	76.8	482	9	AW629814
C 42	19.2	76.8	487	9	AW160553
C 43	19.2	76.8	510	10	BG515497
C 44	19.2	76.8	515	10	BE393833
C 45	19.2	76.8	530	9	AI690580

ALIGNMENTS

RESULT 1

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB666889 260 bp mRNA linear EST 26-OCT-2001
BB668889 RIKEN full-length enriched, 2 days pregnant adult female
ovary Mus musculus cdna clone E330035012 5', mRNA sequence.

BB666889
BB666889.1 GI:16500522

EST.
house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 260)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Sakai, C., Saito, R., Saito, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, O., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.